



Probabilistic modeling of prints at the microscopic scale

Quoc Thong Nguyen

► To cite this version:

Quoc Thong Nguyen. Probabilistic modeling of prints at the microscopic scale. Signal and Image processing. Université Lille1 - Sciences et Technologies 2015. English. NNT: . tel-01337565

HAL Id: tel-01337565

<https://hal.science/tel-01337565>

Submitted on 28 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial| 4.0 International License

N° d'ordre : 41727

UNIVERSITÉ LILLE 1

THÈSE

pour obtenir le grade de

DOCTEUR

En

Spécialité : **Automatique, Génie informatique, Traitement du signal et images**

présentée et soutenue publiquement

par

Quoc Thong NGUYEN

DOCTORAT DELIVRE PAR L'UNIVERSITÉ LILLE 1

Titre:

Modélisation probabiliste d'impression à l'échelle micrométrique

Probabilistic modeling of prints at the microscopic scale

Soutenue le 18 mai 2015 devant le jury d'examen:

Jury

M. Olivier Alata, Prof. à l'Université Jean Monnet,	<i>Rapporteur</i>
M. Stéphane Derrode, Prof. à l'Ecole Centrale de Lyon,	<i>Rapporteur</i>
M. Patrick Bas, Chercheur CNRS,	<i>Examineur</i>
M. François Cayre, Maître de conférences à Grenoble INP,	<i>Examineur</i>
M. Zbigniew Sagan, Ingénieur à Advanced Track & Trace,	<i>Examineur invité</i>
M. Yves Delignon, Prof. à Télécom Lille,	<i>Directeur de thèse</i>
M. Lionel Chagas, Ingénieur de recherche à Grenoble INP-Pagora,	<i>Co-Encadrant de thèse</i>

Thèse préparée dans le Centre de Recherche en Informatique, Signal et Automatique de Lille
(**CRISAL**)

dans le cadre de l'École Doctorale **SPI 072 (Lille I, Lille III, Artois, ULCO, UVHC,
EC Lille)**

PRES Université Lille Nord-de-France

Probabilistic modeling of prints at the microscopic scale

Quoc Thong NGUYEN

May 18, 2015

*A ma famille,
à mes amis*

Contents

1. Introduction	7
2. Related works & Fundamental background	15
2.1. Printing Technologies and Document image degradation	15
2.1.1. Offset printing	15
2.1.2. Electrophotography	17
2.1.3. Inkjet	17
2.1.4. Document image degradation	18
2.2. Printing modeling	19
2.3. Fundamental backgrounds	22
2.3.1. Methods of Estimation	22
2.4. Conclusion	31
3. Experimental analysis	33
3.1. Materials and set-up	33
3.1.1. Materials	33
3.1.2. Pattern analysis	34
3.1.3. Parameters of a 2D object	38
3.2. Analysis	38
3.2.1. Offset printing	39
3.2.2. Electrophotography printing	47
3.3. Conclusion	49
4. Microscopic printing Probabilistic-based model	55
4.1. Model construction	55
4.2. Ponctual parametric modeling	56
4.2.1. Theoretical analysis by moments and histogram	56
4.2.2. Histograms fitting	60
4.3. Conclusion	66
5. Spatial binary model	69
5.1. Binary response model	69
5.2. The Gaussian power kernel	71
5.3. Estimation methods	76
5.3.1. Maximum Likelihood estimation	76
5.3.2. Bayesian estimation method	79
5.4. Experimental results and performance	81
5.4.1. Maximum likelihood estimation	81
5.4.2. Bayesian estimation method	87

Contents

5.4.3. Estimation result from printed dots	89
5.5. Conclusion	90
6. Markov spatial binary model	93
6.1. Interaction model with Markov chain	93
6.2. Estimation	95
6.3. Experimental results and performance	97
6.3.1. Numerical performance with simulated data	98
6.3.2. Estimation result from printed dots	101
6.4. Conclusion	102
7. Conclusions	105
Bibliography	109
A. Material	119
A.1. Parametric distribution	119
A.2. Dot gain	120
B. Calculation in detail	121
B.1. Calculation of Fisher information (5.8)	121
B.2. Calculation of Fisher Information Matrix of model (5.17)	122
B.3. Limit of the KL distance (5.19)	123
B.4. Another real dots	124

Acknowledgments

I would like to take this opportunity to say thank you to all the persons who have contributed in different aspects for this dissertation. First of all, I would like to express my deepest appreciation to my “directeur the thèse” Mr. Yves Delignon, Professor at Telecom Lille, for his pedagogic guidance, constructive criticism, patience and support during my PhD life. I feel fortunate to work with him, he gave many useful ideas from our interesting discussion. Without his persistent help I do not know where this work would have been. I would not forget my co-advisor Dr. Lionel Chagas, at Grenoble INP-Pagora, thanks for his profession with his valuable advices concerning technical issues. Specially for the time in Grenoble, Lionel, Nadège and Brigitte have helped me a lot with the experimental part.

I want to express my sincere thanks to Prof. Olivier Alata, Prof. Stéphane Derrode, Dr. François Cayre, Dr. Patrick Bas, Mr. Zbigniew Sagan, as committee members, for spending their precious time, energy to review this thesis and to attend my defense. Each member had invested a lot of care on my work and given many valuable comments and suggestions.

I am grateful to the secretary of the laboratory Cristal, Ms. Chistine Yvoz, for aiding me with the administrative procedures. A particular thanks goes to the staff at Telecom Lille and at Grenoble INP for helping me with the technical problems. I also want to send a heartfelt thanks to Dr. Wadih Sawaya for his support, not only in the academic works but also for his encouragement to me.

A special thanks is directed to Prof. Nguyen Bac Van, Prof. Duong Minh Duc who supported and encouraged us to pursue an advance education. To my master advisor, Dr. Jing-Rebecca Li, for my time at CMAP.

I would not forget to mention the friends I am lucky to have, Le Thu, Anh Thu, Bao Ân, Nghia, Huy, Hung, Tuan, Maëlle, Romain, Sophie, Oussama, Karthik, Mirabelle, Noura, Xia, Rim, Mehdi ... and other friends that made my PhD life in Europe unforgettable, sincerely thank you.

Last but not least, I would like to express my deep gratitude to my beloved family, my parents (Dao Thi Su, Nguyen Huu To), my brothers (Thach, Trung). Thank you for your love, your supports, it is my fortune to have you in my life.

Abstract

In this dissertation, we develop the probabilistic models of the print at the microscopic scale. Among them, we study the shape randomness of the dots that originates the prints, and the new models should improve many applications such as the authentication in subsequent works, printer identification, or other spatial distribution problems.

Before the modeling, an analysis has been conducted on various supports and various printers. This study shows a large variety of shape of the dots that depends on both the printing technology (conventional offset, waterless offset, inkjet and laser) and the printing support (coated or uncoated).

After analyzing the randomness of the microscopic prints, the digital scan of the microscopic print is modeled in two parts: the gray scale distribution and the spatial binary process modeling the printed/blank spatial distribution.

For the gray scale model, we seek the best parametric distribution that takes into account the specific distributions of the blank and printed areas for any printer or technology. Parametric distributions are selected from a set of distributions with shapes close to the histograms and with the Kolmogorov-Smirnov divergence.

The spatial binary model takes into account the wide diversity of dot shape and, as a consequence, the range of variation of spatial density of inked particles. As a first model, we propose a field of independent and non-stationary Bernoulli variables whose parameters form a Gaussian power kernel. The second spatial binary model encompasses, in addition to the first model, the spatial dependence of the inked area through a inhomogeneous Markov model.

Due to the complexity of the models, two iterative estimation methods are developed; a quasi-Newton algorithm which approaches the maximum likelihood and the Metropolis-Hasting within Gibbs algorithm that approximates the minimum mean square error estimator.

The performances of the algorithms are evaluated and compared on simulated images. The accuracy of the models is analyzed on a set of microscopic scale printings coming from conventional offset, waterless offset, inkjet and laser printers. Results show both the good behavior of the estimators and the consistency of the model with the diversity of micro scale prints.

Résumé

Dans cette thèse, nous développons des modèles probabilistes pour l'impression à l'échelle micrométrique. Tenant compte de l'aléa de la forme des points qui composent les impressions, les modèles proposés pourront être ultérieurement exploités dans différentes applications dont l'authentification de documents imprimés ou de biens de consommation ou encore permettront l'identification d'imprimante ou autres problèmes de répartition spatiale.

Préalablement à la modélisation, une analyse de l'impression sur différents supports papier et par différentes imprimantes a été effectuée. Cette étude montre que la grande variété de forme des points imprimés dépend de la technologie de l'imprimante (jet d'encre, laser, offset waterless, offset conventionnel) et du support d'impression (couché ou non couché).

Le modèle proposé tient compte à la fois de la distribution du niveau de gris et de la répartition spatiale de l'encre sur le papier.

Concernant le niveau de gris, les modèles paramétriques des surfaces encrées et vierges sont obtenues en sélectionnant les distributions paramétriques dans un ensemble de lois de forme similaire aux histogrammes et à l'aide de la divergence de Kolmogorov-Smirnov comme critère de sélection.

Le modèle de répartition spatiale de l'encre est binaire. Devant tenir compte à la fois de la variabilité de forme des points à l'échelle micrométrique et de la variation de la densité spatiale des particules d'encre, le premier modèle consiste en un champ de variables indépendantes de Bernoulli non stationnaire dont les paramètres forment un noyau gaussien généralisé. Un second modèle de répartition spatiale des particules d'encre est proposé, il tient compte, en plus de la décroissance spatiale de la densité de particules encrées, de la dépendance des pixels à l'aide d'un modèle Markovien non stationnaire.

Dûs à la complexité des modèles, deux méthodes d'estimation itératives ont été développées, l'une approchant le maximum de vraisemblance par un algorithme de Quasi Newton et la seconde approchant le critère de l'erreur quadratique moyenne minimale par l'algorithme de Metropolis Hasting within Gibbs.

Les performances des estimateurs sont évaluées et comparées sur des images simulées. La précision de la modélisation est analysée sur des jeux d'images d'impression à l'échelle micrométrique obtenues par des rotatives offset conventionnel, offset waterless, et imprimantes laser et jet d'encre. Les résultats montrent le bon comportement des estimateurs et l'adéquation du modèle à une grande variété d'impressions à l'échelle micrométrique.

Chapter 1.

Introduction

Along with the booming of the global trade, the growth of forgery and counterfeit goods are inevitable. Therefore, Intellectual Property Rights (IPR) have been established to protect exclusive rights to creations and to fostering innovation. However, counterfeiting is like an increasing plague, according to European Commission [1]: *“In 2011, over 114 million articles suspected of infringing intellectual property rights were stopped by customs at the EU external border. The estimated value of the equivalent genuine products is over 1,2 billion euro.”* This is really a headache not only to the manufacturers but also to the consumer who is directly damaged from the counterfeit products. For instance, the public health is highly endangered by counterfeit pharmaceutical products [2]. Even worst, this dirty money can be used to supply the terrorist organizations or other illegal activities. And many other major social problems are caused by forgeries and counterfeit goods. As a consequence, fighting against forgery and counterfeit products becomes essential.

A lot of innovative technologies have been designed to increase the security of the authenticated process. The authentication technologies can be split into three levels [3]:

1. First level: the authentication features are apparent, they can be detected directly without specific devices. For instance, special coatings that changes the color according to the viewing angle [4].
2. Second level: some simple equipments are required such as UV lamp or an inexpensive microscope. For example, a print contains material that is visible under UV light [5].
3. Third level: forensic features are extremely covert. The security system involves expert and sophisticated equipments. They may use an imperceptible tag that requires specialized and secret test to read it. This level of authentication is valuable in case the counterfeit products and packaging are very similar to the genuine articles. For example, the authentication features are extracted from the random structure of paper texture as an intrinsic “fingerprint” [6, 7, 8], such a system is linked to high definition data stored in the database. Fig. 1.1 shows the microscopic texture of the paper surface.

Let us take an example with paper banknote, at the first level, public customers are able to spot forgeries by directly feeling the paper or/and verifying special symbols or portraits embedded on the banknote. At level two, normally proceeded by

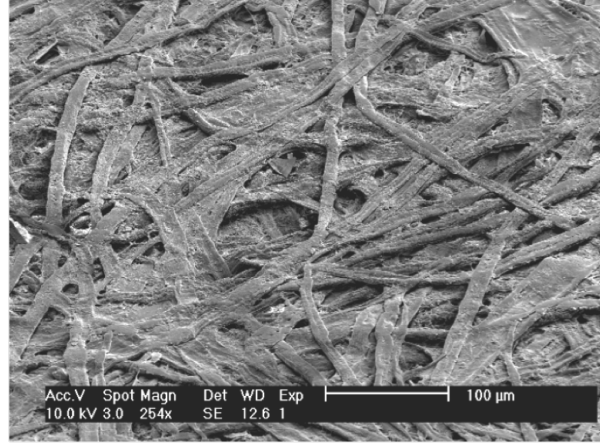


Figure 1.1.: Paper surface under microscopic level [9].

personnel, a simple device such as UV scanner can be required. A UV stripe embedded in the banknote is visible with a back light and can be illuminated when exposed under a UV light. At the forensic level, the optical detection of fiber structure is combined with the digital signature based on public key codes to protect the banknote [10]. Print design is one way of embedding the security feature which is the focus of this work.

The thesis is a part of project Estampille funded by The French National Research Agency (ANR). The goal of the project is to fight against forged printed documents and counterfeit products by inserting Graphical Codes on the documents or the packaging of the good. Anti-counterfeiting method based on graphical code is very popular among others methods because it is cheap and easy to implement in production line. The general framework of the project is depicted in Fig. 1.2:

- A simulated GC is generated from a random source and model the legitimate printing channel to print the GC out, called printed original GC, it is then inserted into the legal product.
- Once the printed on a package is authenticated, the opponent observes the degraded GC, and tries to process it in order to print it by his printing channel to create a reprinted forged GC.
- Both printed original and reprinted forged GC are observed by the receiver, and in order to detect the fake product, the receiver has to post-process (or not) these GCs and then to perform the authentication test.

The graphical code enables to perform both integrity check of printed documents and authentication. Moreover, the authentication process is practical, even with a smartphone. For example, a secure visual recognition system developed by a partner of the project, Advanced Track & Trace®, is found on all wine bottles of Cru Bourgeois du Medoc [11]. The graphical code can be printed with high resolution up to 2400 dpi. With the very small modules (down to $10.58 \mu m$, theoretically) corresponding to the smallest printable dots at high resolution (up to 2400 dpi), this code has higher data storage density and is nearly uncopiable. Fig. 1.3 illustrates a

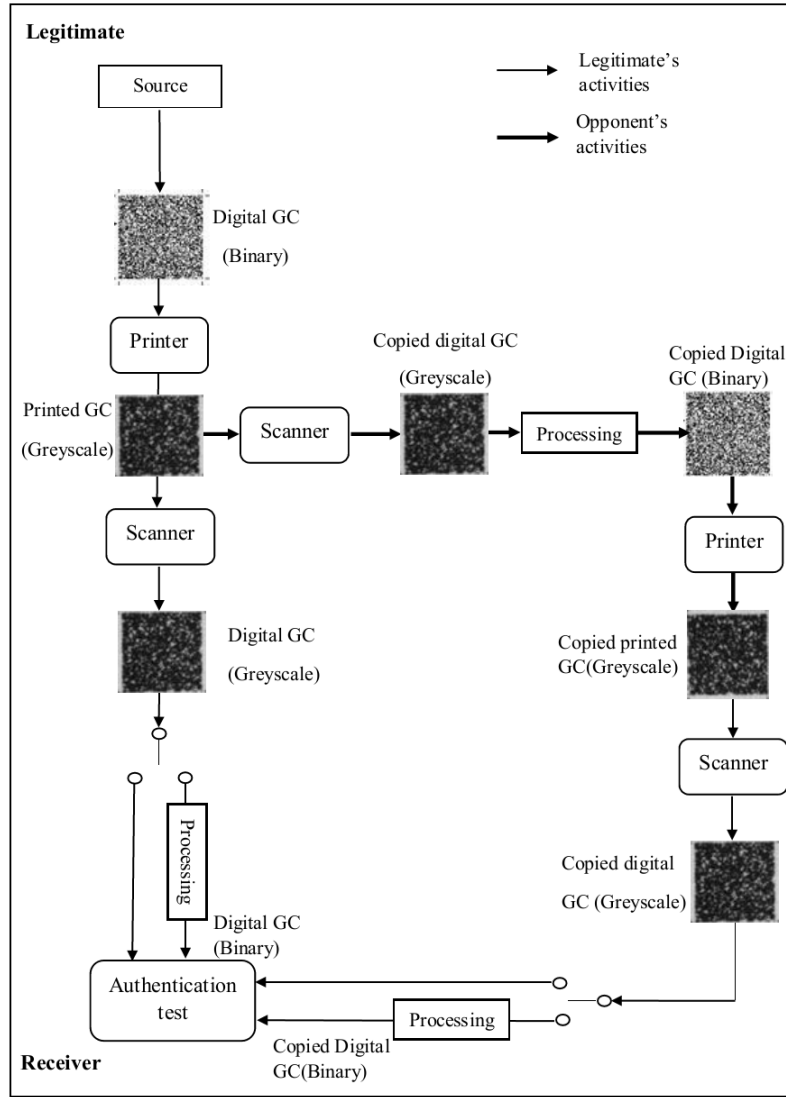


Figure 1.2.: Principle of authentication using graphical codes (GC).

simulated binary code, printed authentic code, and printed fake code. The tasks of our team at LAGIS include :

1. Using information theory for design code and authentication. This approach was aimed by my colleague Anh Thu Phan Ho and her advisers. They study the authentication model under two settings, namely non-channel coding based authentication and channel coding based authentication. While my other colleague Bao An Mai Hoang and his advisers focused on hypothesis testing to improve authentication, and studied the impacts of parameters estimation to the performance of document authentication [12, 13, 14, 15].
2. The second approach is to study the process of printing, this task is taken into account by myself and my advisers. In this study, the prints are analyzed and modeled under physical and signal processing perspective. The print-and-scan model is interesting in authentication using graphical codes, since it is required to perform hypothesis testing. Moreover, by modeling accurately the model,

one is able to maximize the authentication performance [13, 14]. The digital scan of microscopic ink dots after the printing process has to be characterized in order to quantify the distortion of the prints.

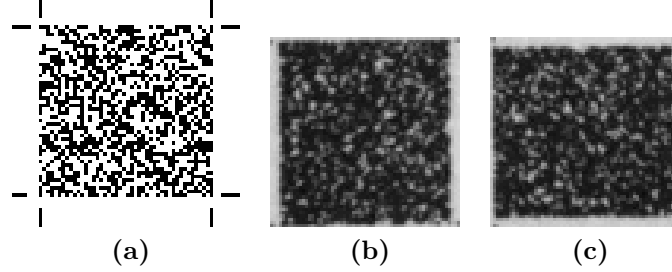


Figure 1.3.: Simulated binary code (a), printed GC (b), copied printed GC (c).

The microscopic details of printings are often unnoticeable under human vision, but there is still a substantial difference between them. Many halftoning-based printer models developed take into account the optical illusion and the varying size of ink dots to create different tones of gray color, and to improve the quality of the images [16, 17, 18, 19, 20, 21, 22]. Other studies focus on image degradation caused by printing and scanning [23, 24, 25], and try to enhance the optical character recognition accuracy rate. Nevertheless, these studies do not consider the characteristics of the microscopic dots. Norris and Smith [26] proposed a probabilistic model in which the ink dots of electrophotographic printing process created by the toners are modeled. However, the stochastic behavior of the microscopic prints has not been deeply analyzed so far.

Under microscopic scale, the characteristics of the print are no longer deterministic but highly random. When the ink is applied on the substrates, the complex interaction between ink and the micro-structure of the substrate causing the random shape of the print is unpredictable. The study of the random structure of the print at the microscopic scale is till an open problem. So how does this problem interest our project? As stated, the graphical code at small scale (printed at high resolution) can carry more information and is difficult to be copied. But when the code is designed, the procedure to identify it also has to be constructed. Before approaching that level, the spatial spreading phenomena of the ink has to be interpreted, and the random shape of the print has to be modeled.

In this thesis, we physically interpret the spatial spreading phenomenon of the ink on the substrate according to principle of the printing process and to the structure of the material. We propose and develop a probabilistic model to simulate the shape of the ink dots under the microscopic scale, the parameters of the model characterizing the physical and optical distortion of the ink spreading. The proposed model is based on the observation of the samples from various technologies and materials. Fig. 1.4 is a dot printed by a laser printer. This model can simulate the stochastic formation of the ink dots as well as the physical interaction between them. By manipulating the parameters, the model is able to simulate the various shape of digital scans of

these microscopic dots from various technologies or printers. An early application of the model is to characterize the random shape of an individual printed dot that can be applied to classify the technologies [27], or determine a specific printer [28].

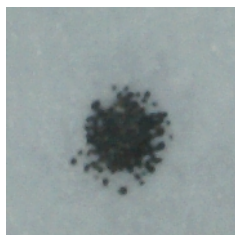


Figure 1.4.: Printed single dot by a laser printer under resolution 600 dpi.

Moreover, the parameters estimation of the model is constructed solidly. This is a challenging problem due to the distinction of the model. The accuracy of the estimator of each parameter is also analyzed. Two main statistical estimation approaches are applied to this model.

- One is executed under the frequentist point of view, in which the parameters are considered deterministic, and the most common method (maximum likelihood estimation) has been developed. Since it is impossible to derive the estimators analytically, numerical methods have been applied.
- The second approach is Bayesian inference, in which the parameters are random. One of the most popular method is *minimum mean square error* which is concerned in calculating an integral. When the model is nonlinear and not Gaussian, the most popular methods for these problem are Monte Carlo methods.

The structure of this thesis is organized as follow

Chapter 2

To give an overview about the printing process, the commonly used printing technologies and the associated principle process are also briefly presented. Some proposed printing models as well as their properties are briefly reviewed. The application of each model is described, and the advantages and disadvantages of each models are also mentioned. In addition, the parameter estimation methods used throughout the thesis are detailed. In particular, the problem of maximum likelihood estimation with optimization under constraints is recalled; other numerical methods such as Markov Chain Monte Carlo estimation are also described.

Chapter 3

To propose an accurate model of the printing process, we need to understand the profile of a single dot at the microscopic scale and the interaction between dots. In chapter 3, the samples of microscopic prints of several patterns from different printing technologies and materials are collected to examine. The effects of each printing technologies on the formation of the micro-patterns are examined. The

explanation for the physical degradation caused by printing processes and material is also given in the chapter. The parameters of the geometrical features such as area, perimeter, are used to characterize the images. The behavior of these parameters is interpreted by the physical properties of the printing processes and the materials.

Chapter 4

This chapter addresses the issue of modeling both the gray level and the shape of printed dot at the microscopic scale. A global probabilistic model for the digital gray image is proposed to fulfill the properties obtained from the analysis of dots conducted in chapter 3. The model includes two parts, the gray-scale model for the histograms of the blank and printed areas, and the spatial spreading of the binary image at the microscopic scale. This chapter also focuses on finding a set of parametric models for the gray-scale level of the blank and printed areas based on four printing technologies and two kinds of paper. The gray-scale model is important because it can be used for hypothesis testing and for channel optimization. We want to understand how consistent the model is on different printers using the same technology. Finally, we study the temporary stability of the parametric estimation when the print is produced by the same printer for a period of time, three weeks in our case.

Chapter 5

In this chapter, the second part of the global model introduced in chapter 4, the spatial binary model, is studied. At the microscopic scale, the dots are made of many inked particles potentially unconnected and the shape is random. We analyze the properties of the real dot to propose a probabilistic model that encompasses various shapes of microscopic prints. Since a good model also required an accurate estimation, an estimation procedure to approximate the parameters of the model is established. The maximum likelihood estimation and Metropolis-Hasting within Gibbs algorithm are both developed for the parameters approximation. The errors of the estimation are analyzed based on the simulated data and the model is tested with the real images from four printing technologies.

Chapter 6

The model proposed in this chapter can be considered as an evolution of the model in chapter 5. This model aims to enhance the accuracy with respect to real prints. Based on the experimental observation in chapter 3, the accuracy of the model can be improved. We propose a way using the Markov random field to take into account the influence of the vicinity to compute the probability of getting a pixel black. So we construct a model to fulfill both the non-stationarity of images of dots and to be coherent with the properties obtained from the analysis of dots conducted in chapter 3. The parameters estimation based on MCMC method, Metropolis-Hasting within Gibbs algorithm, is developed. The accuracy and efficiency of the estimation are analyzed with the simulated data. We also test the model with real images and compare to the model in chapter 5.

Chapter 7

The final chapter is used to summarize our conclusion and outline the possible approaches for the future researches.

Chapter 2.

Related works & Fundamental background

In this chapter, the commonly used printing technologies and their processes are briefly presented. Image degradation is described as well. We also review some proposed printing models taking into account image degradation and their specific applications. The advantages of each model and their limits are also exposed. In the latter, we describe the estimation methods that are used in the dissertation. Particularly, frequentist methods such as maximum likelihood estimation are recalled, we also mention about the numerical methods for solving equations. Bayesian based Monte Carlo methods with a powerful approach like Monte Carlo Markov Chain is also recalled.

2.1. Printing Technologies and Document image degradation

Printing is a reproduction process which applies printing ink to a printing substrate (e.g. a sheet of paper) in a specific mean to store the information (graphics, text). In the experiment, we utilize offset printing technology which can be divided into two popular systems: conventional offset and waterless offset printing technology. We do an experiment on electrophotographic printer. Other commonly used printing technologie is inkjet.

2.1.1. Offset printing

This is a lithographic technology [29], based on the physical phenomena of the immiscible liquids. In this technology, the printing areas are ink-accepting whereas the non-printing parts are ink-repellent. The inked image is transferred from a printing plate, which contains the form of the image, to a rubber blanket, then to the printing surface. The principle process is shown in Fig.2.1. There are two common systems to obtain an ink-repellent areas:

- *Conventional offset printing:* The image-carrier plate is dampened with dampening solution (water with additives). Only the non-printing areas are water-receptive (hydrophilic), whereas the image areas are almost totally unreceptive

to solution. Afterward, the next rollers cover the printing parts with ink delivered from the ink duct, the very fine film of dampening solution preventing the transfer of ink.

- *Waterless offset printing*: This process eliminates the water or dampening system used in conventional printing. Instead of using water-based solution, the printing plate is coated with a highly ink-repellent silicone on the ink-free areas. The printing ink is only accepted on the areas in which the silicone coating has been removed. Therefore, waterless offset offers really shortened make-ready times and simplifies the printing process. Fig. 2.2 shows the printing plates of conventional offset and waterless offset printing.

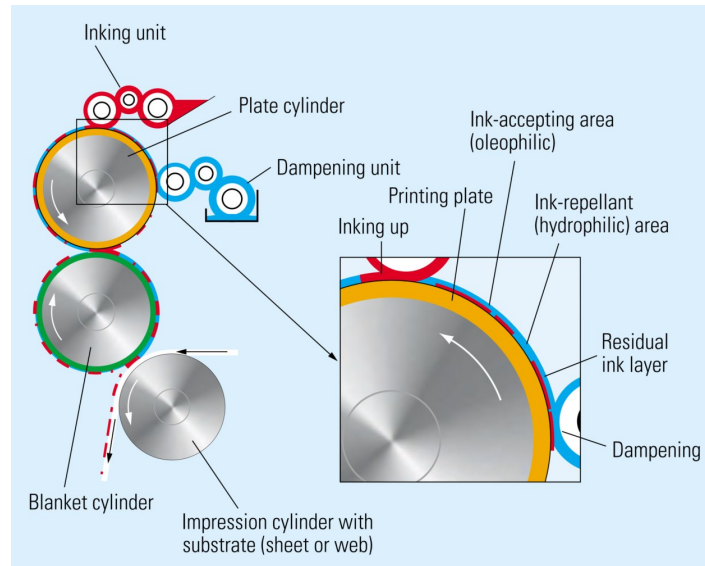
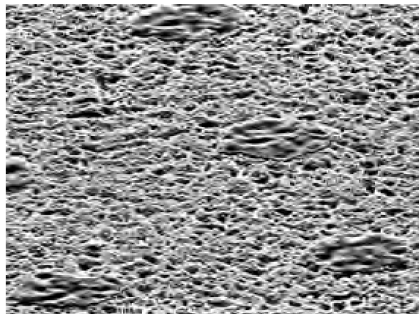
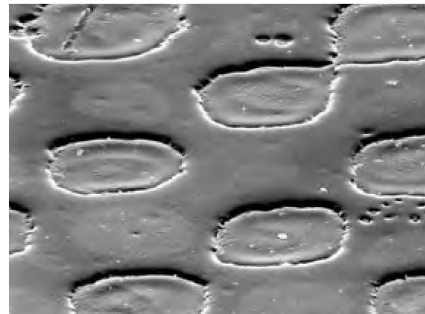


Figure 2.1.: Schematic diagram of offset printing (Fig. 1.3-18 [29])



(a) Aluminum plate of conventional offset



(b) Silicone coated plate of waterless offset

Figure 2.2.: Offset printing plates (Fig. 1.3-20 [29]).

2.1.2. Electrophotography

This is also called laser printing technology. In this process, the laser beam is used to charge a suitable photoconductive surface on the drum with the image or text. Afterward, the charged drum collects the special inks, powder or liquid toners, which have the opposite charge with the photoconductor surface. This ink-carrier drum, then, transfers these toners to the printing substrate, these ink particles are transferred to the paper by the contact pressure between the drum and the paper surface. The paper sheet is then conveyed through two rollers; by heat and pressure, they create a stable image on the paper. Finally, individual particles of toner are cleaned by brush or/and suction, while residual charged areas are neutralized by homogenous illumination, then the drum is charged again for the new printing cycle. There are a number of patents linked to this technology [30, 31, 32, 33]. The principle process is illustrated in Fig. 2.3.

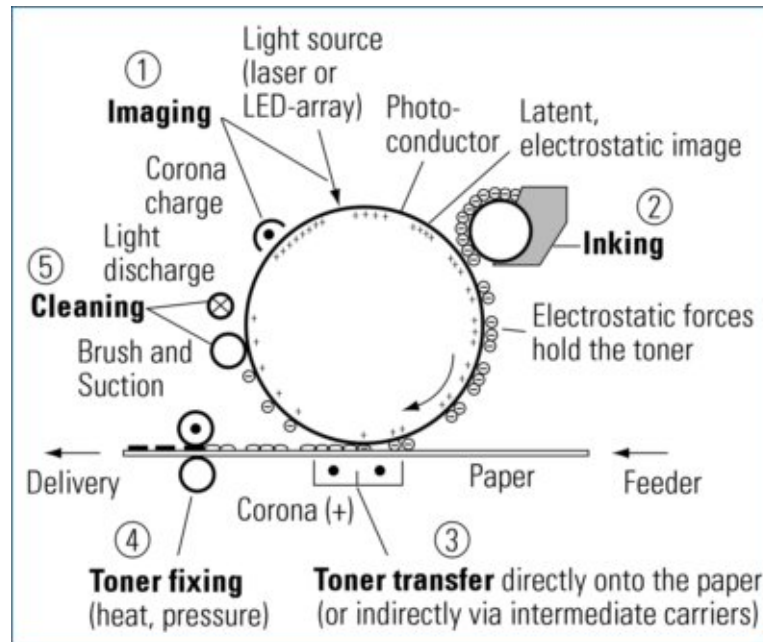


Figure 2.3.: Electrophotography printing process (Fig. 1.3-27 [29])

2.1.3. Inkjet

In this technology, the ink is transferred directly onto the paper. There are *continuous* ink jet and *drop on demand* inkjet. In the *continuous* inkjet technology, a stream of ink is generated constantly, the mechanical phenomena by which a liquid stream disintegrates was studied by Lord Rayleigh [34], these droplets are charged according to the image. These charged droplets are deflected when they fly through the magnetic field. According to the imaging signal, the droplets are directed either onto the paper or back to the system. In this technique, the large part of the stream is driven back into the ink tank.

With the *drop on demand* technology, the ink is only released when the printing part comes. Thermal ink jet and piezo ink jet printing are the most important *drop*

on demand technologies [29]. In the thermal ink jet, they use bubble which contains heat sensitive liquid; according to the imaging signal, the ink drops are squeezed out when the volume of the bubble is expanded by heat. With piezo ink jet, the ink drop is catapulted out of the nozzle when the imaging signal arrives and makes the jet chamber deform. These ink jet technologies are illustrated in Fig. 2.4.

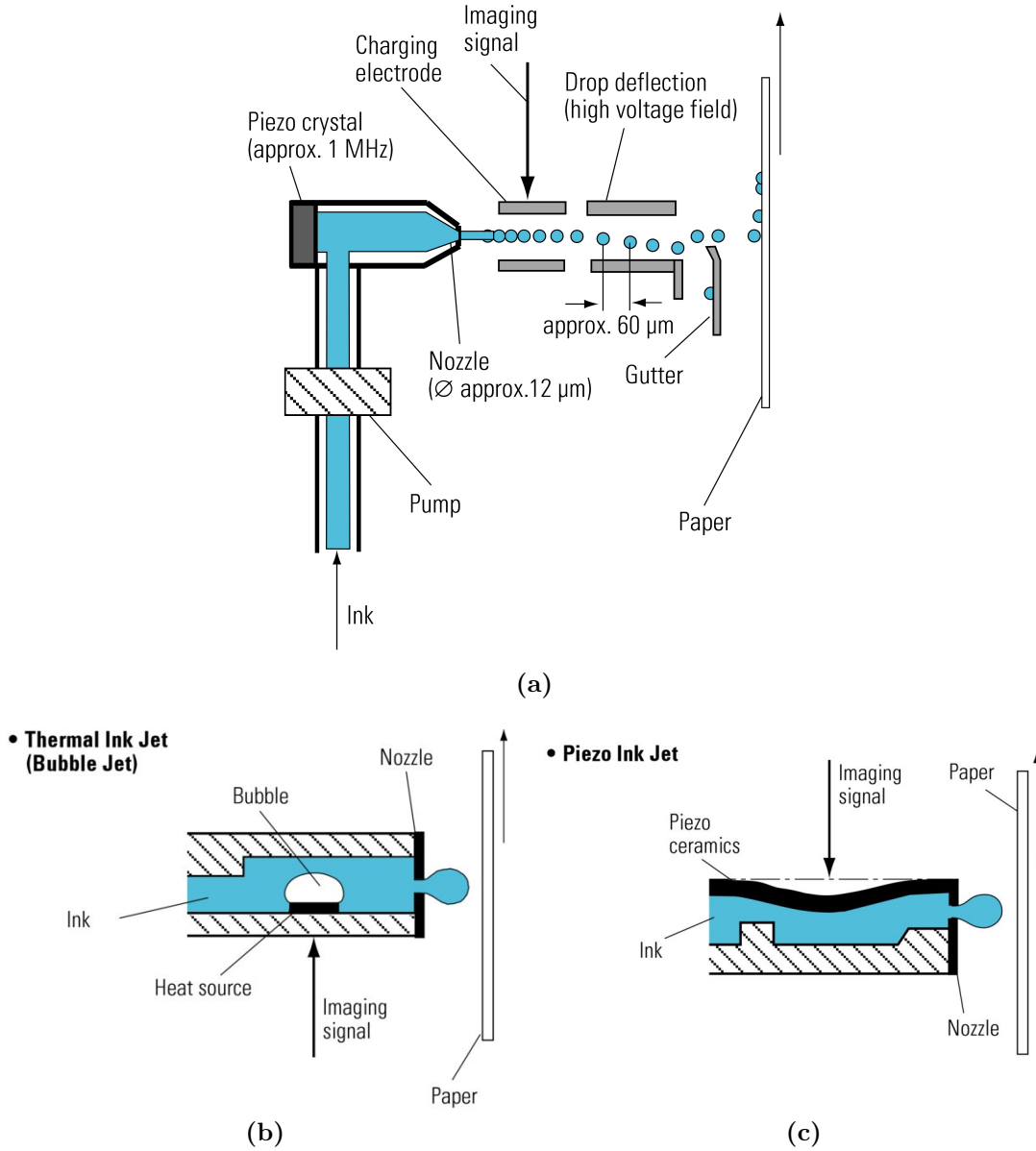


Figure 2.4.: Ink jet technologies, continuous ink jet (a) and drop on demand ink jet (b,c)(Fig. 1.3-30 [29])

2.1.4. Document image degradation

Image degradation is an inevitable issue in the course of printing, reproducing, scanning, capturing to/from a physical document image, even when it is negligible

to human eyes. This loss of quality, even slightly, is mainly responsible for the inaccuracy of the text/image recognition systems. A great diversity of mechanical reasons degrading the text/image is reviewed in [35]:

- variations of paper positions (skew, translation, etc);
- flaking and weak adhesion of ink/toner to printed material; in particular, weak adhesion maybe a problem for laser printing;
- low print contrast, this phenomena is noticeable between conventional offset print and waterless offset print;
- smudging of the paper surface;
- spatter, scatter, and spreading of ink/toner;
- Dot gain includes mechanical and optical dot gain, detailed information about dot gain can be found in sec. A.2.

2.2. Printing modeling

There are a number of printing models proposed in years. Depending on the purposes of the applications such as in optical character recognition (OCR) or in print authentication, a specific model is considered in the scale of the experiments/applications. One of the main purpose of a printing model is to accurately simulate the printed document before the printing process, which may improve significantly the OCR systems or other document analysis issues. For instance, halftoning technique is a method that takes advantage of the optical illusion of human vision to display continuous-tone gray-scale image with only black or/and white dots [36, 37]. The fundamental idea is to induce variation in size, in shape or in spacing of the dots, which generate a gray like effect of the image (see Fig. 2.5). The scale of the model is the resolution of the printer. Model-based halftoning was developed to have richer gray tone as well as to obtain better quality images [16, 17, 18, 19, 20, 21, 22].

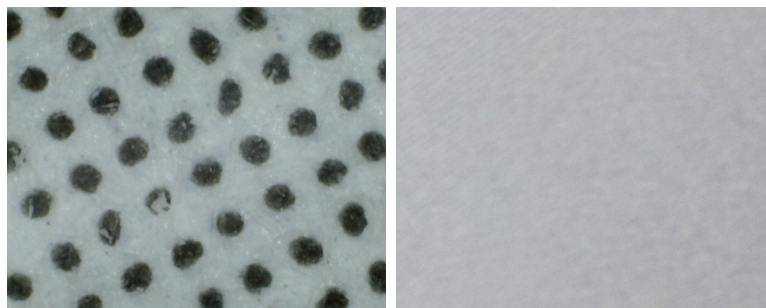


Figure 2.5.: Left: halftone dots. Right: gray color under human vision from a sufficient distance.

In [17, 21], Pappas *et al.* proposed a printer model to digital halftoning for standard laser printers. Digital halftoning is a process of transforming a gray array

into a binary array before being transmitted to the printer, the printed image then has a continuous-tone under the optical illusion of human vision. This model models the dots as circles when printed, so-called “Circular Dot-Overlap Model”, and assigns a constant gray value for the pixels. Given that 1 stands for black pixels while white pixels are 0, given a binary array $[b_{i,j}]$ with size $N_W \times N_H$ that specifies the printed dots, and an array $[p_{i,j}]$ that denotes the gray values having the same dimension, $N_W \times N_H$, as binary array, the circular dot-overlap model assigns the values of gray array $[p_{i,j}]$ as following:

$$p_{i,j} = \mathcal{P}(W_{i,j}) = \begin{cases} 1, & \text{if } b_{i,j} = 1 \\ f_1\alpha + f_2\beta - f_3\gamma, & \text{if } b_{i,j} = 0 \end{cases} \quad (2.1)$$

where $W_{i,j}$ consists of $b_{i,j}$ and its eight neighbor pixels

$$W_{i,j} = \begin{bmatrix} b_{nw} & b_n & b_{ne} \\ b_w & b_{i,j} & b_e \\ b_{sw} & b_s & b_{se} \end{bmatrix},$$

where f_1 is the number of black dots within its four nearest neighbors (i.e. $\{b_n, b_w, b_e, b_s\}$), f_2 is the number of diagonally neighboring black dots that do not have any horizontally or vertically neighboring black dots (i.e. $b_{nw} = 1$ and $b_n = b_w = 0$), and f_3 is the number of pairs of a horizontal neighboring black dot and a vertical neighboring black dot (i.e. $b_s = b_e = 1$). α, β, γ are the ratio determined by a specific resolution of the printer and the detailed expression can be found in the papers [17, 21]. This model is quite accurate to many printers but not to all [18]. Moreover, the gray level of each pixel is assumed to be constant instead of being assigned in a continuous spatial domain. Another drawback of this model is that it does not work for the very high resolution printers such as 2400 dpi or 1200 dpi.

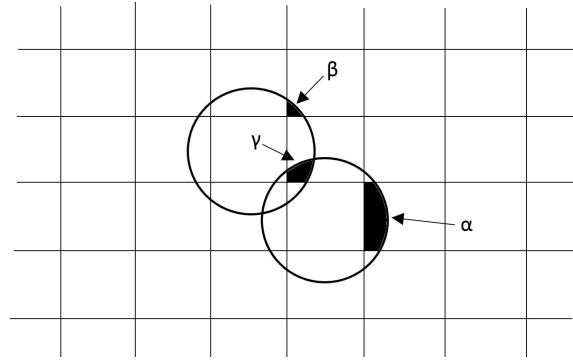


Figure 2.6.: Definition of α, β , and γ for model (2.1)

A printer model not only contains the properties of a printing process but also includes image degradation which is inevitable in the course of printing, photocopying, and scanning. The models of document image degradation have been discussed in [23, 38]. In [35], document image degradation also takes into account the characteristics of the printing process. This degradation model is based on either physics

of printing and imaging or statistical surface of image distribution.

Henry S. Baird, in [23], had proposed a physics-based model, in which the parameters defining imaging defects are described:

- Resolution: The image is digitized (in pixels) under a specific resolution (in pixels/inch).
- Blur: The spreading of a point is modeled as a circularly symmetric Gaussian filter. This phenomena is caused when the image is out of focus.
- Threshold: Binarization the gray-scale image.
- Sensitivity: Distribution of per-pixel additive noise.
- Jitter: Random dislocation of the pixels sensor centers from the ideal square grid.
- Skew: Rotation of the document.
- Width, height: scaling factors, horizontally and vertically.
- Baseline: Relative position between the text and the conventional baseline.

The parameters of this model are considered as random variables, they are trained by a large number of sample before being used to simulate the defect model or apply to a pattern recognition systems.

In [38, 39, 40, 41], Tapas Kanungo *et al.* proposed a model for the optical distortion that occurs while scanning or photocopying a thick, bound document. This strategy models the degradation of the document at the bended part. Moreover, the idea for local distortion was demonstrated under probabilistic point of view. The probability of a pixel changing from its value, black or white, is represented as a function of the distance of that pixel from the boundary of a character. Let d be the distance of a foreground or background pixel from the boundary of the character and α, β are the scale parameters. The probabilities $P(1 \mid d, \beta, f)$ and $P(0 \mid d, \beta, f)$ represent the chance of a foreground pixel at a distance d to remain as 1 or to switch to 0, respectively. Similarly, $P(1 \mid d, \alpha, b)$ and $P(0 \mid d, \alpha, b)$ are the probabilities, from a distance d , of a background pixel to change to 1 or to remain at value 0.

Under the microscopic printing, Margaret Norris and Elisa H. Barley Smith proposed a probabilistic model for the average coverage of the toner particles ink of electrophotographic printing process [26]. A single dot printed from a laser printer is a cluster of toner particles, with this model, each toner particle is considered as an ideal circle characterized by r . With the assumption that these N toner particles are dispersed on the paper proportionally to the energy of the laser trace, the Gaussian density function is used with spreading parameter σ . The region of the paper is discretized in the square units, the centers of the toners are restricted in these unit squares. The probability that a unit square is covered by at least one toner particle is

$$P(\text{square is black}) = 1 - P(\text{no circle covers the unit square}) \quad (2.2)$$

Assuming that each particle has the same size, shape, and independent to each other. So

$$P(\text{square is black}) = 1 - (1 - P(C))^N \quad (2.3)$$

where $P(C)$ is the probability that the square pixel is covered by one toner particle. The model is validated with a visible appearance, Fig. 2.7 illustrates the simulation. However, the model is only valid for electrophotographic (laser) printing process, and the estimation of the parameter has not been considered yet.

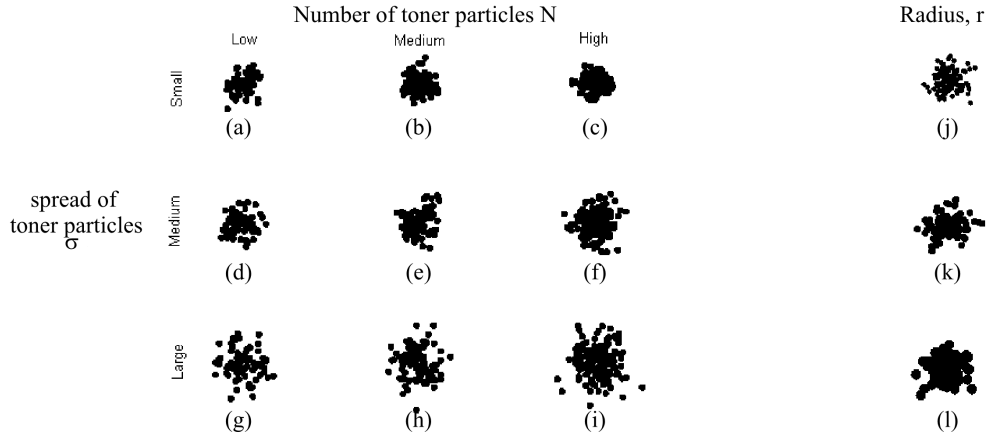


Figure 2.7.: Simulation of a single dot with $\sigma \in \{0.35, 0.50, 0.71\}$ and $N \in \{50, 100, 150\}$ (a-i), and different toner particles radii $r \in \{0.14, 0.20, 0.28\}$ (Fig. 3 in [26])

2.3. Fundamental backgrounds

2.3.1. Methods of Estimation

Frequentist parameter estimation

Frequentist inference draws the distribution of a sample data by emphasizing the frequency of the observed data. In this part, we briefly describe the methods utilized through the thesis. **Method of moments** is a classical and simple approach in which one derives an equation or system of equations of the empirical moments and the probabilistic moments which depend on the parameters of interest. The estimated parameter is defined as the solution (if exist) of the equation(s). Another popular method is **Maximum-likelihood estimation**.

The idea of maximum likelihood estimation method is to find the parameter $\hat{\theta}$ in Θ that is most plausible with the observed data \mathbf{x} . The likelihood function has the same form with the density function but differs in meaning. The joint density function for all observations is $f(x_1, x_2, \dots, x_n | \theta)$. Under different perspective, the variable is θ while the observation is fixed, it is called the likelihood function

$$L(\theta; x_1, x_2, \dots, x_n) = f(x_1, x_2, \dots, x_n | \theta) \quad (2.4)$$

In practice, the logarithmic form is often more convenient to handle, $\mathcal{L}(\theta; \mathbf{x}) = \log L(\theta; \mathbf{x})$. The maximum likelihood estimator (MLE) is obtained as

$$\hat{\theta}_{ML} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta; \mathbf{x}). \quad (2.5)$$

If the log-likelihood function, or likelihood function for simplicity, is differentiable in Θ° , the interior of Θ , the possible MLE's are the values $\theta \in \Theta^\circ$ satisfying

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0. \quad (2.6)$$

Since the sample size increases to infinity, under some conditions, the maximum likelihood estimators are consistent, i.e. the sequence of MLE's converges in probability to the true value [42], $\hat{\theta}_{ML} \rightarrow_p \theta_0$. In addition, with some regular conditions, the estimators have asymptotically normal distribution (Theorem 3.3 in [42]), $\sqrt{n}(\hat{\theta}_{ML} - \theta_0) \rightarrow_d \mathcal{N}(0, I^{-1})$, with I the Fisher information matrix. Moreover, this estimation is asymptotically efficient, which means the covariance matrix of the estimators achieves the Cramer-Rao lower bound when the sample size tends to infinity.

In practice, the estimator is supposed to be in Θ° and the likelihood is twice differentiable. The solution is found by solving (2.6), then verified by the second derivative of the log-likelihood \mathcal{L} . In many applications, the solution of (2.6) might not be obtained analytically, then a numerical method is necessary.

Quasi-Newton methods

Without loss of generality, the problem we discuss is a minimization problem. These methods are based on the Newton's method in optimization. It enables to avoid both complex calculation of the Hessian matrix and ill-conditioned matrix problems. The Newton's direction is derived from second-order Taylor expansion of the real function f

$$f(x_k + p) \approx f(x_k) + p^T \nabla f_k + \frac{1}{2} p^T B_k p, \quad (2.7)$$

assuming the Hessian matrix B_k is positive definite, then it is desirable to find the vector p that minimize the right-hand side of (2.7), denoted by $m_k(p)$. By setting the gradient of $m_k(p)$ to zero, we obtain the Newton's step

$$p_k = -B_k^{-1} \nabla f_k. \quad (2.8)$$

In practice, we usually have a numerical issue in approximating the inverse of Hessian matrix. One of the most popular formula for updating $H_k := B_k^{-1}$ is the Broyden-Fletcher-Golfarb-Shanno (BFGS) method [43], H_0 is initialized by the identity matrix I , and is updated by

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k} \right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k} \right) + \frac{s_k s_k^T}{y_k^T s_k}, \quad (2.9)$$

where

$$s_k = x_{k+1} - x_k, \quad y_k = \nabla f_{k+1} - \nabla f_k.$$

Moreover, to attain a sufficient decrease in the objective function f , a positive scalar α_k is used to adjust the step length (2.8). This scalar satisfies the Wolfe conditions [44, 43]

$$\begin{aligned} f(x_k + \alpha_k p_k) &\leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k, \\ \left| \nabla f(x_k + \alpha_k p_k)^T p_k \right| &\leq c_2 \left| \nabla f_k^T p_k \right|, \end{aligned} \quad (2.10)$$

where $0 < c_1 < c_2 < 1$, (2.10) is called *strong Wolfe conditions*. Then the quasi-Newton algorithm proceeds as in Algorithm 2.1.

Algorithm 2.1 Quasi-Newton algorithm with BFGS method.

Initialize $x_0, H_0 = I$

Choose ϵ

while $\|\nabla f_{k+1} - \nabla f_k\| > \epsilon$ **do**

$s_k = -\alpha_k H_k \nabla f_k$, α_k satisfies (2.10)

$x_{k+1} = x_k + s_k$

 Compute ∇f_{k+1}

 Update H_{k+1} with H_k, s_k and $y_k = \nabla f_{k+1} - \nabla f_k$ using (2.9)

end while

Optimization under constraints

Constraints do frequently appear in optimization problems. Therefore, efficient numerical algorithms that have ability to approximate the solution with the constraints are interesting. The constrained problem is stated as follow

$$\min_x f(x), \quad (2.11)$$

subject to:

$$\begin{aligned} g_i(x) &\leq 0, \quad i \in \mathcal{I}, \\ h_j(x) &= 0, \quad j \in \mathcal{J}. \end{aligned}$$

We shall briefly describe the augmented Lagrangian method [45, 46] used to approach the problem, further reading can be found in [47] or Chapter 17 in [43]. In order to transform into the non-constrained problem, the augmented Lagrangian function is introduced:

$$L_A(x, \mu, \lambda, r) = f(x) + \sum_i \mu_i g_i(x) + \sum_j \lambda_j h_j(x) + \frac{r}{2} \left(\sum_i \max\{0, g_i(x)\}^2 + \sum_j h_j(x)^2 \right). \quad (2.12)$$

The quadratic forms in (2.12) are the penalty functions which vanish when the constraints are satisfied and “penalize” the Lagrangian multiplier when the constraints are violated at the current point x . As r increases severely, the minimizer of non-constrained problem (2.12) approaches the solution of the constrained problem (2.11). Without loss of generality, we consider only one equality constraint and one inequality constraint. The optimal solution of the problem (2.11) satisfies (see Theorem 12.1 [43], [48]):

$$\nabla f(x) + \mu \nabla g(x) + \lambda \nabla h(x) = 0, \quad (2.13)$$

for appropriate λ and $\mu \geq 0$. Assuming that at iteration k -th, we have approximated $\mu^{(k)}$, $\lambda^{(k)}$ and $x^{(k)}$, it is desirable to improve the approximation of the multipliers $\mu^{(k+1)}$, $\lambda^{(k+1)}$. The optimality condition for (2.12) with $\mu^{(k)}$, $\lambda^{(k)}$ and some $r^{(k)}$ is

$$\nabla f(x) + \mu^{(k)} \nabla g(x) + r^{(k)} \max \{0, g(x)\} \nabla g(x) + \lambda^{(k)} \nabla h(x) + r^{(k)} h(x) \nabla h(x) = 0, \quad (2.14)$$

and the approximation $x^{(k)}$ is expected to be close to the true optimal solution x . Therefore, if we compare (2.13) and (2.14), the appropriate update on the multipliers is

$$\begin{aligned} \lambda^{(k+1)} &= \lambda^{(k)} + r^{(k)} h(x^{(k)}), \\ \mu^{(k+1)} &= \max \{0, \mu^{(k)} + r^{(k)} g(x^{(k)})\}. \end{aligned}$$

Recall that g_i , h_j and f are twice differentiable, so that the typical algorithms which take into account the gradient values, such as Quasi-Newton methods, can be applied to the non-constrained problem. The procedure of the algorithm proceeds as follows:

Algorithm 2.2 BFGS algorithm.

Initialization: $\mu_i^{(0)} \geq 0$, $\lambda_j^{(0)}, r^{(0)} > 0$, $c > 1$, and small $\epsilon > 0$

New approximation: Minimizing the augmented Lagrangian function $L_A(x, \mu^{(k)}, \lambda^{(k)}, r^{(k)})$ with respect to $x^{(k)}$

Stopping criterion:

$$\begin{aligned} \left| \nabla f(x^{(k)}) + \mu_i^{(k)} \nabla g_i(x^{(k)}) + \lambda_j^{(k)} \nabla h_j(x^{(k)}) \right| &< \epsilon, \\ \left| \mu_i^{(k)} g_i(x^{(k)}) \right| &< \epsilon, \\ \left| h_j(x^{(k)}) \right| &< \epsilon \end{aligned}$$

Update:

$$\begin{aligned} \mu_i^{(k+1)} &= \max \{0, \mu_i^{(k)} + r^{(k)} g_i(x^{(k)})\} \\ \lambda_j^{(k+1)} &= \lambda_j^{(k)} + r^{(k)} h_j(x^{(k)}) \\ r^{(k+1)} &= cr^{(k)} \end{aligned}$$

where c is usually chosen large to boost the penalty functions.

Fisher information and Cramér-Rao Bound

In mathematical statistics, the Fisher information is a measure of the amount of information that an observable random process carries about unknown parameters upon which it depends. Intuitively, the more information about unknown parameters is contained in an observed data set, the more we can expect to find an accurate estimator. This result has been formalized by Cramér, Rao, Fréchet and Darmois. They showed that the covariance matrix of an unbiased estimator is bounded by the inverse of the Fisher information matrix. As a consequence, the so-called Cramér-Rao lower bound serves as a benchmark to assess the performances of the estimators of the deterministic parameters [49, 50, 51]. Let $\mathbf{T}(U)$ be the unbiased estimator of θ , the Cramér-Rao bound states that the covariance matrix of $\mathbf{T}(U)$ satisfies

$$\text{cov}_\theta(\mathbf{T}(U)) \geq I(\theta)^{-1}, \quad (2.15)$$

where $I(\theta)$ is the Fisher information matrix. And the ordered relation between these two matrices is in the meaning that the matrix $M = \text{cov}_\theta(\mathbf{T}(U))I(\theta) - I_N$ is positive semi-definite, where I_N is an identity matrix. Since M is a Hermitian matrix, then M is positive semi-definite if and only if its eigenvalues are non-negative. If one is only interested in a component of θ , say θ_m , from the property of nonnegative-definite matrix, we have

$$\text{var}(\theta_m) = [\text{cov}_\theta(\mathbf{T}(U))]_{m,m} \geq [I(\theta)^{-1}]_{m,m}. \quad (2.16)$$

Let us note that in information geometry, the Fisher information matrix is used as a metric which is a particular Riemannian metric on a smooth statistical manifold. In that space, points are probability measures and can be used in many applications requiring statistical inference algorithms or machine learning methods.

Bayesian Methods

In the Bayesian paradigm, the prior information that is determined by a prior density $\pi(\theta)$ is combined with the sample information \mathbf{x} and summarized in the posterior distribution $\pi(\theta | \mathbf{x})$. According to Bayes formula

$$\pi(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta)\pi(\theta)}{m(\mathbf{x})}, \quad (2.17)$$

where $m(\mathbf{x}) = \int f(\mathbf{x} | \theta)\pi(\theta)d\theta$ is the marginal unconditional density of \mathbf{X} . While the prior distribution $\pi(\theta)$ reflects beliefs about θ prior to experimentation, $\pi(\theta | \mathbf{x})$ interprets the update beliefs about θ posterior to the observed sample \mathbf{x} . In contrast to frequentist inference, the Bayesian estimation is interested in the value of θ that maximizes the posterior probability $\pi(\theta | \mathbf{x})$, called a maximum a posteriori probability (MAP). Since the marginal unconditional density does not depend on θ , then the estimate is

$$\hat{\theta}_{MAP}(\mathbf{x}) = \arg \max_{\theta \in \Theta} \pi(\theta | \mathbf{x}) = \arg \max_{\theta \in \Theta} f(\mathbf{x} | \theta)\pi(\theta).$$

One would see that $f(\mathbf{x} \mid \theta)$ is a likelihood function. Therefore, the MAP method coincides with Fisher's method of Maximum Likelihood Estimation when the prior $\pi(\theta)$ is uniform. Another common criterion requires a loss function $L(\hat{\theta}, \theta)$ which represents the loss by estimating with $\hat{\theta}$ [52]. The Bayesian estimator $\hat{\theta}$ of θ is chosen such that it minimizes the posterior expected loss:

$$\mathbb{E} [L(\hat{\theta}, \theta) \mid \mathbf{x}] = \int L(\hat{\theta}, \theta) \pi(\theta \mid \mathbf{x}) d\theta.$$

With the squared-error loss ($L(\hat{\theta}, \theta) = \|\theta - \hat{\theta}\|^2$), the Bayesian estimator of θ that minimizes the mean squared error is $\hat{\theta}(\mathbf{x}) = \mathbb{E}_\pi [\theta \mid \mathbf{x}]$, i.e.

$$\hat{\theta}(\mathbf{x}) = \int \theta \pi(\theta \mid \mathbf{x}) d\theta. \quad (2.18)$$

However, it is often impossible to achieve the exact computation or a tractable closed form of the integral (2.18) analytically. Among other approaches, Monte Carlo methods [53, 54] offer a numerical mean to approximate this quantity.

Classical Monte Carlo Methods

The objective of the Monte Carlo method [53] is to approximate the integral (2.18). Monte Carlo methods are stochastic methods, in which a number of random samples are generated from a probability distribution. A Monte Carlo method consists of two kinds of statistical problems: providing a procedure to generate samples from a given probability distribution, and providing a validated numerical estimation of the integral from the samples. Let θ be a random vector distributed by a law $\pi(\cdot)$ with expected value μ and $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}\}$ be a sequence of n i.i.d samples drawn from π . Then the Strong Law of Large Number states that the average value converges almost surely to the expected value [55]

$$\frac{\theta^{(1)} + \theta^{(2)} + \dots + \theta^{(n)}}{n} \xrightarrow{a.s.} \mu \quad \text{when } n \rightarrow \infty.$$

This means that for a sufficiently large number of samples of θ , the desired expectation can be approximated by the arithmetic mean

$$\int \theta \pi(\theta) d\theta \approx \frac{1}{n} \sum_{t=1}^n \theta^{(t)}.$$

However, in practice, sampling directly from the distribution $\pi(\cdot)$ is often impossible or the analytical form cannot be derived. Therefore, instead of drawing the i.i.d samples directly from the complicated distribution, one can use easier proposal distribution to generate the samples then apply an adjusted step in order to generate the samples from the target distribution π . There is a variety of sampling methods.

Importance sampling

This is one of the methods that is able to avoid the complexity of the distribution π of interest [56], and it is based on so-call importance function. The method of importance sampling evaluates (2.18) by generating a sample $\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}$ from a given distribution q and by approximating

$$E_{\pi} [\theta] \approx \frac{1}{n} \sum_{t=1}^n \frac{\pi(\theta^{(t)})}{q(\theta^{(t)})} \theta^{(t)}. \quad (2.19)$$

The estimator is based on an alternative representation of (2.18)

$$E_{\pi} [\theta] = \int \theta \pi(\theta) d\theta = \int \theta \frac{\pi(\theta)}{q(\theta)} q(\theta) d\theta = E_q \left[\theta \frac{\pi(\theta)}{q(\theta)} \right]. \quad (2.20)$$

The convergence of (2.19) is guaranteed as long as $\text{supp}(q) \supset \text{supp}(\pi)$, whatever the choice of the distribution q . Actually, $\frac{\pi(\theta^{(t)})}{q(\theta^{(t)})}$ is used to weight the sample. So that the main concern of the method is to design an efficient importance distribution q . Nevertheless, with the high-dimension of θ , the design of a good importance function is really challenging, thus the methods of Markov chain Monte Carlo (MCMC) was proposed.

Markov chain Monte Carlo and Metropolis-Hastings algorithm

MCMC methodology [57] provides a sufficient approach to many complex scientific problems such as in physics, signal processing [58, 59, 60]. MCMC is widely popular due to the profound effect on computation problems in Bayesian statistics.

Definition 2.3.1. For a stationary Markov chain with transition matrix P , a distribution π is called stationary (or invariant distribution) if

$$\pi P = \pi.$$

The idea of MCMC algorithm is to generate a Markov chain having π as its stationary distribution. The Markov chain $(\theta^{(t)})_{t \geq 1}$ which is irreducible and aperiodic has the property of converging to this stationary distribution [61]. Moreover, in the positive recurrent case, we have

$$\frac{1}{n} \sum_{t=1}^n \theta^{(t)} \xrightarrow{a.s.} E_{\pi}(\theta) \quad \text{when } n \rightarrow \infty. \quad (2.21)$$

The convergence of the ergodic average in (2.21) to the desired expectation is guaranteed by the ergodic theorem. The strategy is to run the simulation sufficiently long such that the first m samples can be omitted and the chain begins to draw samples close enough to distribution π . In the other words, $P^{(t)}(\cdot | \theta^{(1)})$ converges to invariant distribution π which does not depend on t or $\theta^{(1)}$. Thus the later dependent samples $\{\theta^{(t)}\}$ are approximately from π [57]. There are a number of papers

discussing this so-called burn-in period and number of the iterations [62, 63, 64, 65]. The expectation then can be approximated with the rewritten ergodic average

$$E_{\pi}(\theta) \approx \frac{1}{n-m} \sum_{t=m+1}^n \theta^{(t)}. \quad (2.22)$$

Metropolis-Hastings algorithm is the most popular MCMC method which allows for construction of the Markov chain [66, 67]. At each iteration t , a candidate point θ^* is drawn from a proposal conditional distribution $q(\cdot | \theta^{(t-1)})$. The candidate point θ^* is accepted to be the next state $\theta^{(t)}$ with probability

$$\alpha(\theta^{(t-1)}, \theta^*) = \min \left\{ \frac{\pi(\theta^*) q(\theta^{(t-1)} | \theta^*)}{\pi(\theta^{(t-1)}) q(\theta^* | \theta^{(t-1)})}, 1 \right\}. \quad (2.23)$$

If the candidate point is not accepted, the chain does not jump, i.e. $\theta^{(t)} = \theta^{(t-1)}$. The algorithm is summarized in Algorithm 2.3

Algorithm 2.3 Metropolis-Hastings algorithm.

Initialize $\theta^{(0)}$

for $t = 1, \dots, n$ **do**

 Generate $\theta^* \sim q(\cdot | \theta^{(t-1)})$

 Compute the acceptance ratio $\alpha(\theta^{(t-1)}, \theta^*)$

 Decide

$$\theta^{(t)} = \begin{cases} \theta^* & \text{with probability } \alpha(\theta^{(t-1)}, \theta^*) \\ \theta^{(t-1)} & \text{with probability } 1 - \alpha(\theta^{(t-1)}, \theta^*) \end{cases}$$

end for

Random walks

For the practical construction of a Markov chain with Metropolis-Hastings algorithm, a natural approach is to take into account the previous value to generate the current value. The common choice is to simulate θ^* according to

$$\theta^* = \theta^{(t)} + \epsilon_t, \quad (2.24)$$

where ϵ_t is a random perturbation with distribution d , independent of $\theta^{(t)}$. Then the conditional distribution $q(\theta^* | \theta^{(t)})$ is equivalent to $d(\theta^* - \theta^{(t)})$. The Markov chain generated as 2.24 is a random walk. The distribution d is usually chosen to be symmetric, $d(\theta^* - \theta^{(t)}) = d(\theta^{(t)} - \theta^*)$, which leads to the acceptance ratio as following:

$$\alpha(\theta^{(t-1)}, \theta^*) = \min \left\{ \frac{\pi(\theta^*)}{\pi(\theta^{(t-1)})}, 1 \right\}. \quad (2.25)$$

The common choices for distribution d are uniform, normal or student's t distributions. When π is a posterior probability, the Metropolis-Hasting algorithm consists in choosing θ^* as soon as the posterior of candidate point θ^* is larger than the posterior of $\theta^{(t-1)}$, otherwise, the candidate point θ^* is chosen with probability $\frac{\pi(\theta^*)}{\pi(\theta^{(t-1)})}$ or the state remains unchanged in next step, $\theta^{(t)} = \theta^{(t-1)}$.

Gibbs sampling algorithm

In practice, when the dimension of θ is large, the exploration in the space is not tractable. In those cases, it is more convenient and computationally efficient to divide the vector of parameters θ for sampling. One of the well-known efficient method proposed is Gibbs sampler [68]. Assuming that θ is partitioned into B sub-components such as $\theta = [\theta_1, \theta_2, \dots, \theta_B]$, and each block is drawn one by one successively. Each block θ_b^* is generated from its conditional distribution given all other block, $\pi(\theta_b | \theta_{-b}^{(i-1)})$, where $\theta_{-b}^{(i)}$ represents the samples of all blocks except for θ_{-b} at iteration i , $\theta_{-b}^{(i-1)} = [\theta_1^{(i-1)}, \dots, \theta_{b-1}^{(i-1)}, \theta_{b+1}^{(i-1)}, \dots, \theta_B^{(i-1)}]$. The Gibbs sampling algorithm proceeds as Algorithm 2.4

Algorithm 2.4 Gibbs sampling algorithm.

```

Initialize  $\theta^{(0)} = [\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_B^{(0)}]$ 
for  $i = 1, \dots, n$  do
  for  $b = 1, \dots, B$  do
    Draw  $\theta_b^{(i)} \sim \pi(\theta_b | \theta_{-b}^{(i-1)})$ 
  end for
end for

```

Metropolis-Hastings within Gibbs

In addition, a combination of Gibbs sampling and Metropolis-Hastings was suggested in [69, 70], called Metropolis-Hastings within Gibbs. This strategy is helpful in the problems when sampling from the full distribution is complicated. The algorithm is summarized in Algorithm 2.5

Monte Carlo Standard Error

The standard error of a Monte Carlo estimate has major statistical meaning. A Markov chain $(\theta^{(t)})_{t \geq 1}$ with invariant distribution π , under regularity conditions [71], the Central Limit Theorem (CLT) is held; that is

$$\sqrt{n}(\bar{\theta}_n - E_\pi(\theta)) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \quad \text{when } n \rightarrow \infty \quad (2.26)$$

where $\bar{\theta}_n = \frac{1}{n} \sum_{t=1}^n \theta^{(t)}$, the ergodic average and σ^2 is the variance. Since the variance is unknown, we can use its estimate. The estimate of the standard deviation is

Algorithm 2.5 Metropolis-Hastings Gibbs sampling algorithm.

 Initialize $\theta^{(0)} = [\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_B^{(0)}]$
for $i = 1, \dots, n$ **do**
for $b = 1, \dots, B$ **do**

 Generate $\theta_b^* \sim q(\theta_b \mid \theta_1^{(i)}, \dots, \theta_{b-1}^{(i)}, \theta_b^{(i-1)}, \theta_{b+1}^{(i-1)}, \dots, \theta_B^{(i-1)})$

Compute the acceptance ratio

$$\alpha(\theta_b^{(i-1)}, \theta_b^*) = \min \left\{ \frac{\pi(\theta_b^* \mid \theta_{-b}^{(i-1)}) q(\theta_b^{(i-1)} \mid \theta_1^{(i)}, \dots, \theta_{b-1}^{(i)}, \theta_b^*, \theta_{b+1}^{(i-1)}, \dots, \theta_B^{(i-1)})}{\pi(\theta_b^{(i-1)} \mid \theta_{-b}^{(i-1)}) q(\theta_b^* \mid \theta_1^{(i)}, \dots, \theta_{b-1}^{(i)}, \theta_b^{(i-1)}, \theta_{b+1}^{(i-1)}, \dots, \theta_B^{(i-1)})}, 1 \right\}$$

Decide

$$\theta_b^{(i)} = \begin{cases} \theta_b^* & \text{with probability } \alpha(\theta_b^{(i-1)}, \theta_b^*) \\ \theta_b^{(i-1)} & \text{otherwise} \end{cases}$$

end for
end for

provide as

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (\theta_i - \bar{\theta}_n)^2}. \quad (2.27)$$

Suppose that for any α , we want to construct the confidence interval of $1 - \alpha$, i.e. find z_α such that $P(-z_\alpha \leq Z \leq z_\alpha) = 1 - \alpha$, where Z is the standardized distribution. Given $\hat{\sigma}^2$ and the size of sample n , the confidence interval is approximated by

$$\left[\bar{\theta}_n - z_\alpha \frac{\hat{\sigma}}{\sqrt{n}}, \bar{\theta}_n + z_\alpha \frac{\hat{\sigma}}{\sqrt{n}} \right]. \quad (2.28)$$

2.4. Conclusion

Through this chapter, we described some existing printer models and the physical properties that inspired these models. The goals of these models are mostly to improve the accuracy of the printing and/or scanning, but the characteristics of the micro printing process are not deeply analyzed. We recalled the maximum likelihood estimation which is the most popular frequentist method and MCMC method which plays a very important role in Bayesian based parameter estimation. These are the estimation methods used throughout this dissertation.

Chapter 3.

Experimental analysis

To propose an accurate model of the printings, we need to analyze the print at the microscopic scale. In this chapter, the influences of the physical properties of ink and substrates on the quality of the document with different patterns are presented. The affect of the resolution of the printing process also needs to be interpreted. The analysis is based on the samples which are from conventional offset, waterless offset and laser printer, on two kinds of substrate, coated paper and uncoated paper with rough surface.

A test form containing 9 dots has been realized. All configurations of a 3×3 matrix, which are 512 different configurations, was created. However, due to the repetition of many patterns as well as the insignificance of some, only few major configurations were retained and analyzed. Two main aims in this experiment include:

1. Profile analysis of a single dot,
2. Degradation analysis for the different configurations.

3.1. Materials and set-up

In this section, we describe the material that is used in the experiment. After, a set of patterns is chosen to analyze. The choice of patterns is based on the interaction of dots, and on the influence of the direction of the printing process if it exists. Finally, we propose the parameters which are analyzed for each experimented pattern.

3.1.1. Materials

There are two substrates being studied: coated paper is a kind of paper coated with one or several layers of a white pigment giving the surface a satin finish and reducing the roughness of the paper, and uncoated paper. During the coating process, the paper web is mechanically routed between rollers under pressure, friction and warmth which makes the paper thinner than the uncoated natural paper, also being used in the experiment. Due to the characteristics of each technologies, offset process will use offset papers, and papers for electrophotography are taken for laser printer. A test is done over 10 samples of each kind of paper in order to characterize their properties (see Tab. 3.1).

(a)

Coated paper	Thickness(μm)	Roughness(μm)
Mean	85	0.9
Std. deviation	1	0.1
Uncoated paper		
Mean	149	5.9
Std. deviation	1	0.1

(b)

Coated paper	Thickness(μm)	Roughness(μm)
Mean	98	2.1
Std. deviation	1	0.15
Uncoated paper		
Mean	103	5.1
Std. deviation	1	0.1

Table 3.1.: Characteristics of coated paper and uncoated paper used in offset process (a) and laser process (b).

In addition, the printing ink also plays a particular role in each technology. For offset printing, the printing ink is generally highly viscous, dynamic viscosity $\eta = 40\text{--}100$ Pa.s (see [29] page 137). Furthermore, for the conventional offset printing, the ink is able to achieve the additive of dampening solution, and may, typically, contains water [72]; on the other hand, silicone oil can be added to the ink of waterless offset printing to guarantee that the non-printed image areas are not inked (see [29] page 137). Waterless offset ink has a comparatively higher viscosity and is generally stiffer than conventional offset ink. Due to the fundamental difference between the two systems, conventional offset and waterless offset prints are different. In the “wet” process, conventional offset printing, the solution has lower cohesion than the ink, the water may remain on the ink film then make lower contrast on the image and create the “brighter” area inside the printing parts (see Fig. 3.2a and Fig. 3.3a). On the other hand, for the electrophotography printing, special ink needs to be applied, called toners. Toners can be powder or liquid, which explains why the image have the shape of a cluster of particles Fig. 3.1. The inks used in ink jet technology are in liquid form, so the printing is very dependent on the substrate surface [29].

3.1.2. Pattern analysis

For a given support, many parameters can influence the surface, the perimeter and the compactness of the printed dots. The inking adjustment is essential regardless of the printing process. It is necessary to calibrate the machine. The first step of the calibration consists in adjusting the ink quantity for the solid color. An optimum optical density (OD) is defined (in connection with the ink thickness). The second step consists in adjusting the different percentages of coverage (5-95%). The luminance gradient from 0 to 95% of coverage should gradually decrease. The

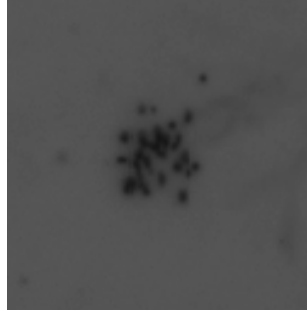


Figure 3.1.: Image from a laser printer on uncoated paper, the image is captured under the magnification of 400 times.

color of a material may be characterized by its reflectance spectrum (the light which is reflected from the material in the visible range for wavelengths from 400 to 700 nm [73]). For one wavelength, the definition of the optical density is

$$OD = \log_{10} \left(\frac{1}{R} \right),$$

where R stands for Reflectance (quantity of light which is reflected by the paper). In our case, the optimum setting has been achieved before using printed matter. Among 512 different configurations, there are fourteen different patterns analyzed Tab.3.2. For an easier following after, we index these patterns by number. The goals that we expect to understand for the patterns choice are:

- The profile of one printed dot
- The direction influence of the printing process
- The degradation for the different configurations
- The behavior of the dot gain to the non-printed parts

Each pattern is printed to analyze three parameters: area of ink coverage, perimeter, and compactness measures of the shape. Moreover, the form of a single dot is also studied. These patterns were printed on two kinds of substrate, coated and uncoated paper. The dot size of these patterns is, theoretically, $10.58 \mu m$, which means 2400 dpi resolutions of the offset printers.

Since we have mentioned about dot gain, beside the mechanical and physical dot gain, there is also the optical dot gain. The optical illusion happens when a mass is in close proximity to one another (see Fig. A.1), [74]. For each pattern and each kind of process, 100 samples are captured by an AxioCam camera installed on a Zeiss Microscope. This camera directly captures the gray-scale level images, then the threshold method is applied. Besides three parameters mentioned, we also compare the area of the real print to the theoretical area of ink coverage. The theoretical size of the area is based on the theoretical size of a single dot, e.g., the theoretical area of one dot under the resolution 2400 dpi is approximately $10.58 \times 10.58 (\mu m^2)$.

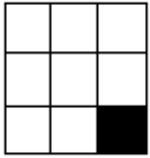
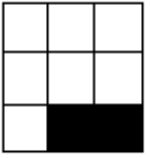
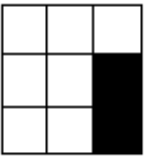
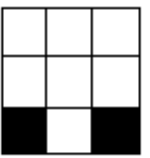
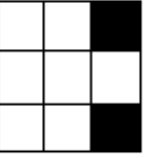
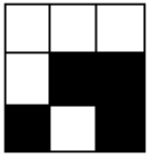
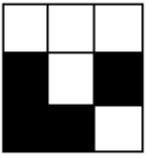
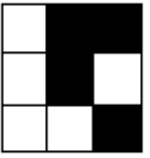
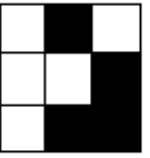
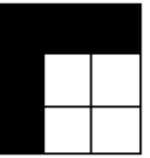
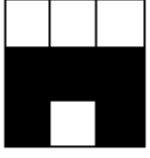
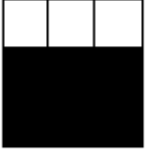


1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.2.: The patterns to observe.

In addition, with these patterns, we also studied the pseudo-resolution of 1200 dpi, which is the group of four dots under 2400 dpi. Therefore, the size of a single dot is, theoretically, about $10.58 \times 2 \times 10.58 \times 2 (\mu m^2)$.

Binary threshold

Under the micro-scale, the printed parts are the main concern in the experiment, then the binary images are more desirable than the gray scale ones. There are a number of threshold methods to convert gray-scale images into binary images [75, 76, 77]. Each method, for a specific purpose, has some advantages and also few drawbacks, based on the requirement of the application. However, in this analysis, we are mainly studying discrimination between printers, materials, technologies. Therefore, it is sufficiently that all the samples are binarized under the same treatment, the maximum entropy method mentioned in Kapur et al. [78] performs properly on these gray images presented in Fig. 3.2 and Fig. 3.3.

Locating dot center

After the binarized step of a single dot, we expect to find the most “theoretical” center of each single dot. At the beginning, the center is simply located by averaging the coordinates of all black pixels which are considered as the part of the dot. However, in this way, we do ignore the error of the printer when it misplaces the true center. Afterward, the borders are considered to identify the center, two horizontal borders and two vertical borders which means the errors caused by the displacement could be reduced.

Fig. 3.4 is the original binary image, after the printing process, each printing dot becomes a cluster. Through the capturing process, each of these dots is roughly

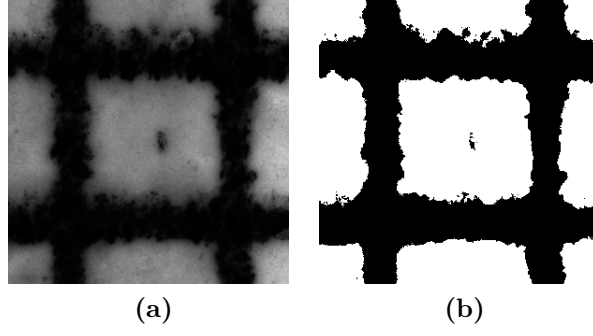


Figure 3.2.: The conventional offset coated print (a.) is binarized by maximum entropy method (b.)

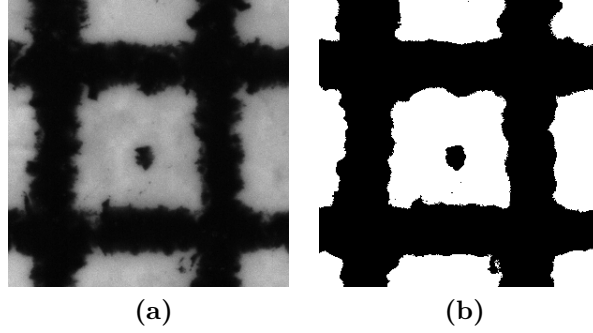


Figure 3.3.: The waterless offset coated print (a) is binarized by maximize entropy method (b).

considered as an ellipse with two diameter k_x, k_y (pixels). We denote s (pixels) the original length of the diameter of that circular shape, then $k_x = S_x/s, k_y = S_y/s$ will become the size of one dot. Therefore, the center dot is located as

$$\begin{aligned} dx &= \text{theoreticaldistance}_x * k_x \\ dy &= \text{theoreticaldistance}_y * k_y \end{aligned}$$

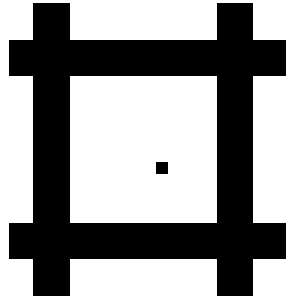


Figure 3.4.: Original binary image.

To estimate the border lines, the simple linear regression is applied, as shown in Fig. 3.5. However, the orthogonal property is also conserved between the lines (see

Fig. 3.5). On the next step, we determine the dot center based on the approximated borders.

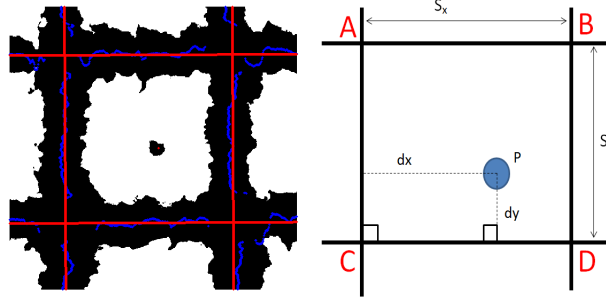


Figure 3.5.: Principle of measurements to determine the dot position.

3.1.3. Parameters of a 2D object

Visually, we distinguish between the conventional offset and the waterless offset process as well as between coated and uncoated paper. However, a procedure for a computer to recognize the origin of these images is more practical. Therefore, the introduction of the following parameters is essential. We consider pixel as the unit of a black and white image. Let us denote I the value matrix of the image with $I(i, j) = 0$ represents the black pixels and $I(i, j) = 1$ is for the white ones. The object X , a dot or a group of dots, will be analyzed through the parameters

- Area, is defined as the number of black pixels, $A(X)$.
- Perimeter, the number of vertices of the black pixels bordering the white pixels, $P(X)$.
- Compactness measure of the shape, the ratio of the area of X to the area of the circle having the same perimeter

$$\Psi(X) = \frac{4\pi A(X)}{P(X)^2} \quad (3.1)$$

this parameter is non-negative and does not exceed 1. The closer to 1 the coefficient is, the more compact the shape is.

- Ratio of experimental area to theoretical area, $R_A(X) = \frac{A(X)}{A_{th}(X)}$.

3.2. Analysis

In this step, we analyze the properties of the printing to distinguish the process as well as the material. According to the law of large number, the average values of the parameters we measure converge when the sample size increases to infinity. This statement indicates that the average of the results are asymptotic to the mean value since the sample size increases (see Fig. 3.6). Moreover, since these samples are i.i.d.

(independent and identically distributed), the Central Limit Theorem states that arithmetic mean converges in distribution to a normal distribution. Due to the limit of the equipment, there is no test form from the ink-jet technology is collected.

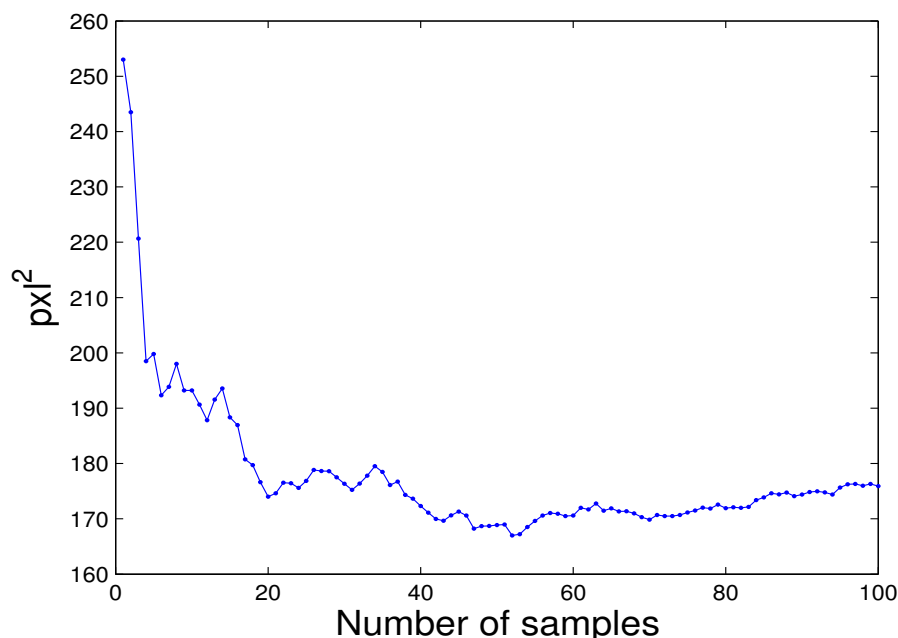


Figure 3.6.: The variation of the average area value from pattern 1 printed by waterless offset on uncoated paper under the resolution 1200 dpi.

3.2.1. Offset printing

In the experiment, the information of two offset printers used in this experiment is given in Tab. 3.3.

Name	CODIMAG	SEAILLE & TISON
Type	Waterless offset	Conventional offset
Web	Semi-rotating system	Rotating system
Width (mm)	420	480
Resolution (dpi)	2400	2400
V-max (m/mn)	50	200
Classical optical density	1.8	1.8

Table 3.3.: The offset printers are used to print the test form.

Patterns under resolution 2400 dpi

The original resolution of the printer, 2400 dpi, is analyzed. Under this high resolution, it is difficult for the printers to create a suitable shape according to the pattern (see Fig.3.7). With the waterless offset process, in general, the coverage

area of the print on the coated paper are greater, visually, than the one on the uncoated paper (see Fig. 3.8a). This phenomena is caused by the physical property on the surface of each paper, the coated paper has a glossy finish making it quite smooth and probably shiny (high gloss). This coating layer restricts the amount of ink that is absorbed, which creates the larger inked coverage on the surface than on the uncoated paper surface. Moreover, due to the roughness of the fiber structure, the perimeter of the print is, generally, greater on the uncoated layer (see Fig. 3.8b), which means that the print on the coated paper is less fragmented. Looking at the variation of the mean values, they are quite similar with the area parameter while with the perimeter ones, the values on the coated paper are more stable. This behavior, under the micro scale, is caused, again, by the nonuniform porous surface of the uncoated paper, some positions are less rough than other. We will also see the effect of these properties on later samples.

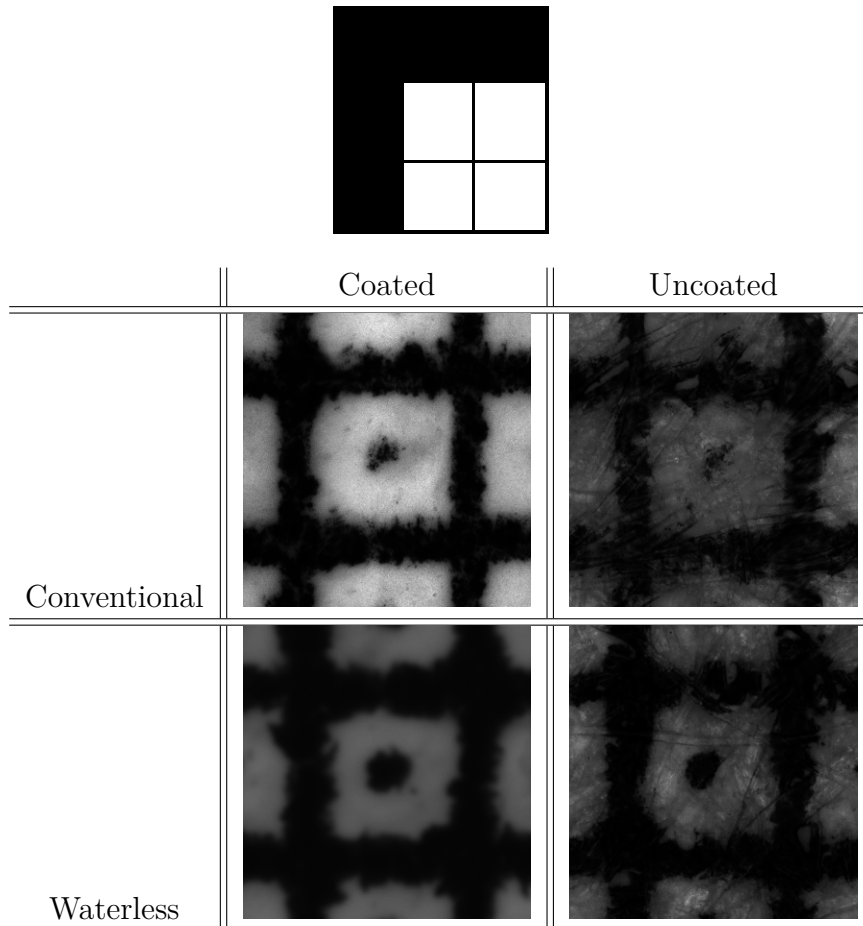


Figure 3.7.: Pattern is printed under the conventional offset and waterless offset process on coated and uncoated paper with the resolution 2400 dpi.

About the conventional process, even on the coated paper, the coverage area is negligible (see Fig. 3.9). It only becomes detectable with the large number of dots. Therefore, the information from the uncoated paper is quite weak for printing the micro image. This result is due to the “wet” process of the technology, as mentioned in sec. 3.1.1. Consequently, with these micro prints, the conventional

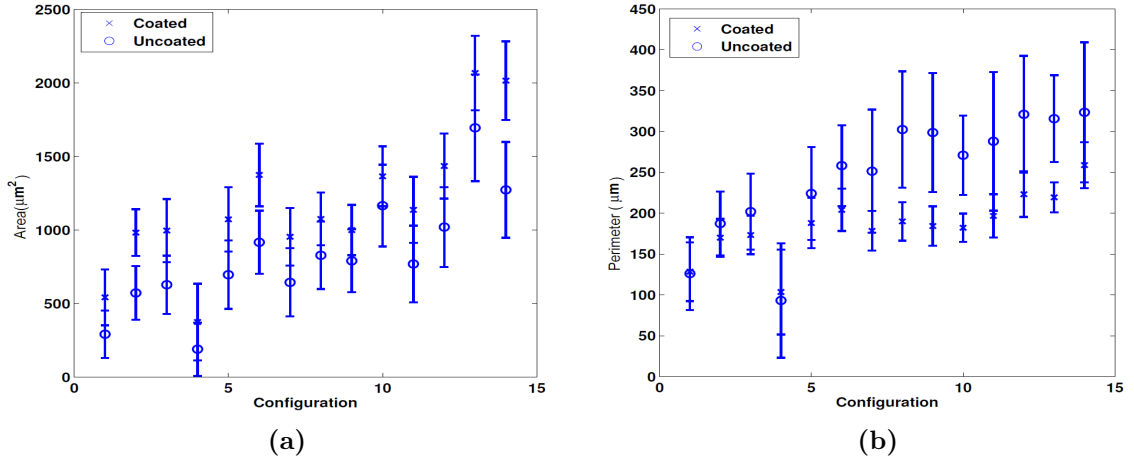


Figure 3.8.: The mean values of area (a.) and perimeter (b.) on both kind of paper with waterless offset process.

technique cannot perform properly. In the following, we will focus on the prints from waterless offset printer.

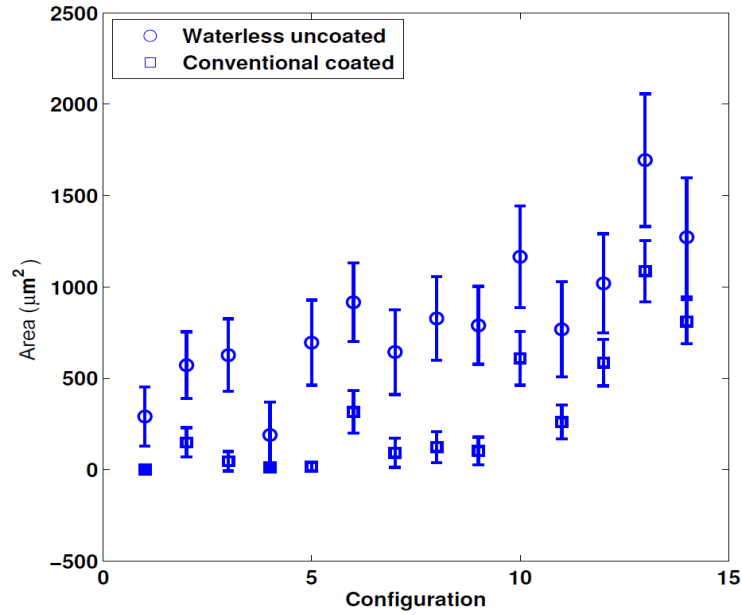


Figure 3.9.: The mean values of the area from the waterless offset with uncoated paper and conventional offset on coated paper.

In these samples, the pattern 4 (see Tab. 3.2) has a different behavior, this pattern is the same as pattern 5 but with different directions of printing. The line crossing two dots in pattern 5 is the direction of the printer. With pattern 4, some images show the missing dot (see Fig. 3.10a,c), where only one of two dots was printed, even in few case there is no printed dot Fig. 3.10b,d. While on pattern 5, two dots were always printed and are mostly connected by the spreading of the ink. This

phenomena causes the small average value of area of pattern 4. This issue happens also on uncoated paper.

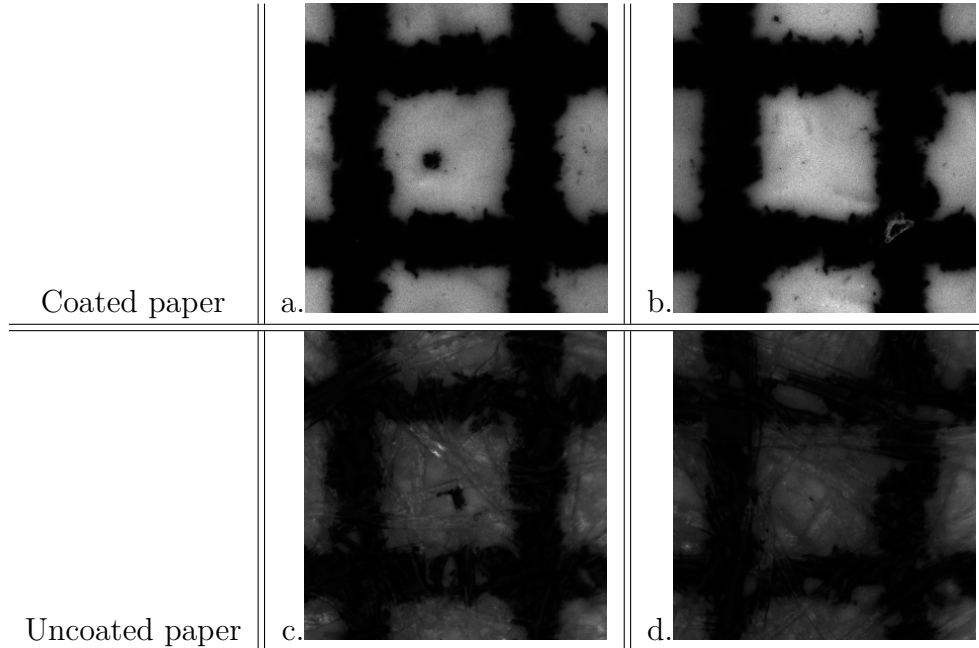


Figure 3.10.: Print of pattern 4 with conventional offset under resolution 2400 dpi, one dot missing (a,c), nothing was printed (b,d)

pattern	Coated paper	Uncoated paper
1	4.81	2.58
2	4.35	2.54
3	4.42	2.78
4	1.66	0.84
5	4.76	3.09
6	3.05	2.03
7	2.11	1.43
8	2.38	1.94
9	2.22	1.75
10	2.42	2.07
11	2.02	1.36
12	2.12	1.51
13	2.29	1.88
14	1.98	1.25

Table 3.4.: Ratio of experimental area to theoretical area of 14 pattern with Waterless offset technology, resolution is 2400 dpi.

Tab. 3.4 shows the difference in area of the real print and the area of the theoretical printing parts. Generally, the more dot patterns, the less area increase, specially when the dots are close to each other. This behavior is not really surprising, if there

is a single dot, the ink spreads to all directions, while with two closed dots, the gain parts overlap the area of its neighbor. The direction of the printing process also causes the difference in areas of the same patterns (patterns 4, 5), the greater area of pattern 5 is due to the trace of the ink when paper moves in the process. This also explains why pattern 6 covers greater area than patterns 7, 8, 9, Fig. 3.11.

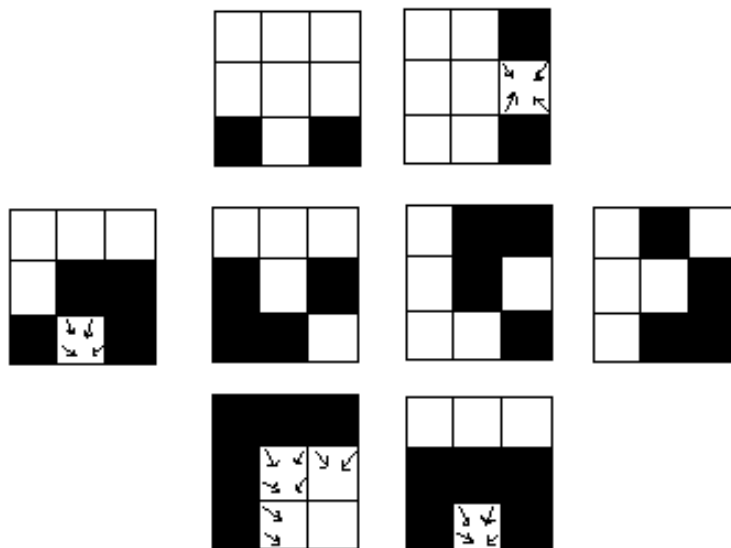


Figure 3.11.: Ink spreading caused by printing direction.

The average images calculated from 100 image of one dot on coated paper is illustrated in Fig.3.12, these figures give an overview of the round shape of the printed dot.

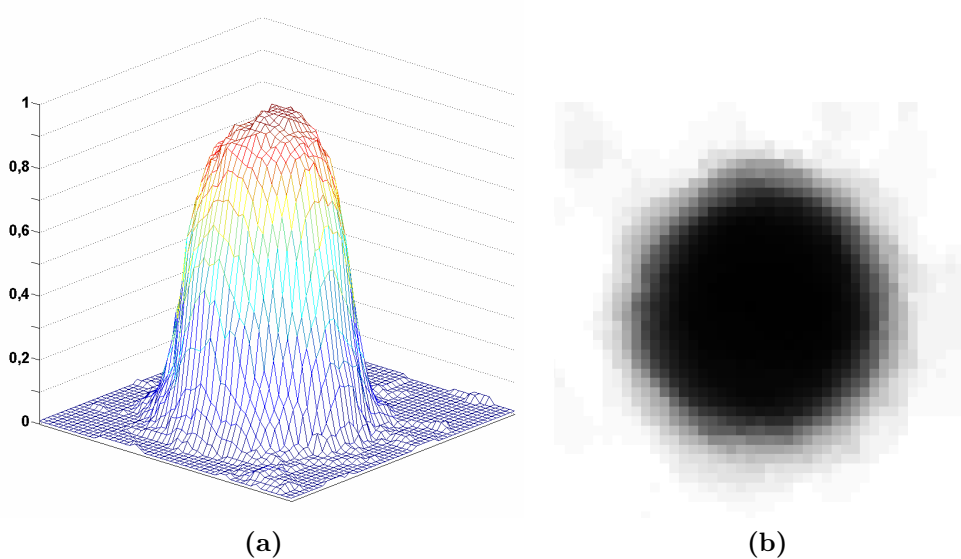


Figure 3.12.: The average image of one dot from waterless offset printer on coated paper under the resolution 2400 dpi.

Based on these observations, the 2400 dpi resolution is too high for the printers to create the regular image, parameters from the samples are rather inconsistent. That is the reason to expect the 1200 dpi solution samples more stable, it is essential for a consistent procedure. Nevertheless, one can conclude that the area values from waterless process and conventional process are distinguishable.

Patterns under resolution 1200 dpi

In this part, the samples under 1200 dpi of these 14 patterns are examined. The mechanical and physical phenomena between the ink and the substrates was explained in sec.3.2.1, so we shall pass these explanations. Considering the waterless offset, the area values are clearly distinguishable between two materials (see Fig. 3.13a). Moreover, whatever the kind of paper, waterless offset always creates larger area pattern and less fragment than conventional technology (see Fig. 3.13c, d). About the conventional technology on two different substrates, the mean values are not significantly distinguishable following with the large variance, the substrates cannot be classified in this case (see Fig. 3.13b). This comment implies that conventional process and waterless process can be discriminated by comparing the coverage area without concerning the substrate.

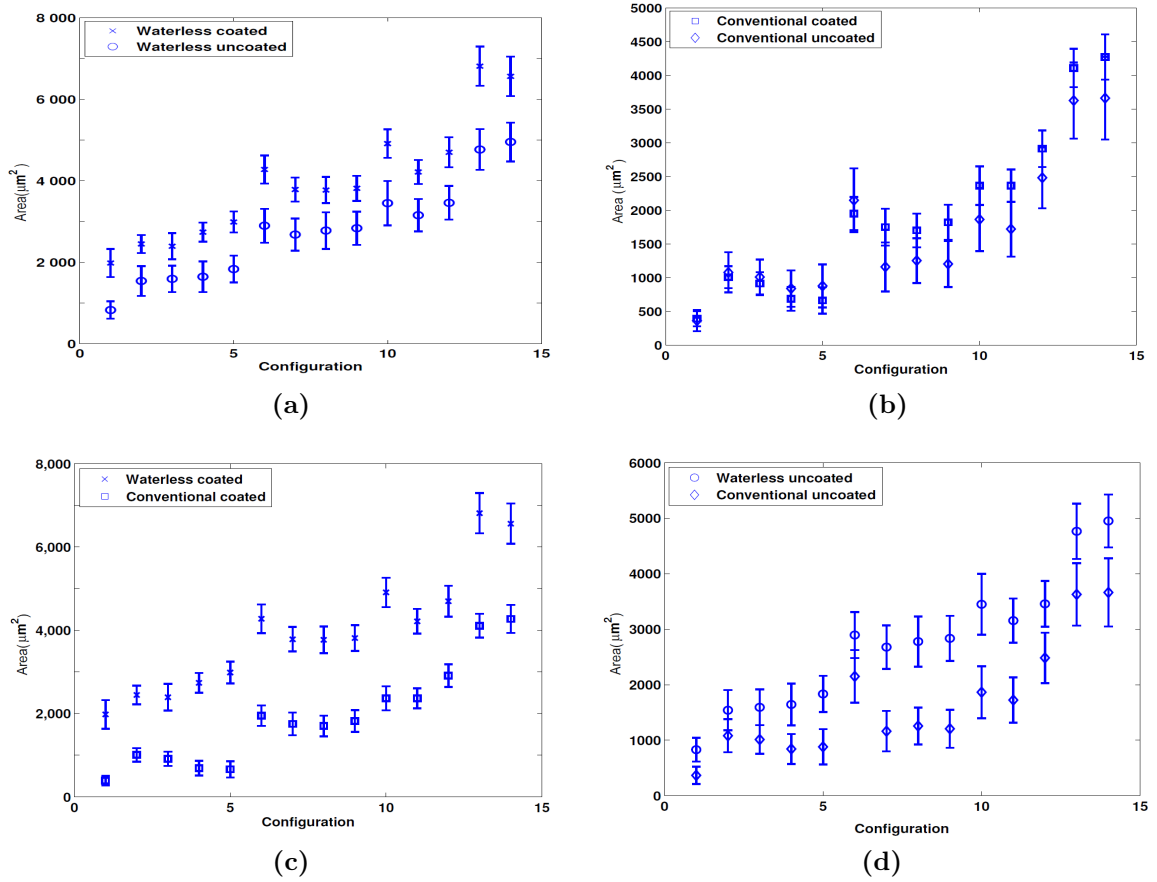


Figure 3.13.: The average values of area, resolution 1200 dpi.

In addition, patterns 4, 5 cover larger area than patterns 2, 3 with waterless offset

printing, area of pattern 5 is greater than area of pattern 4, pattern 6 is also bigger than patterns 7, 8, 9. This phenomena is explained in sec. 3.2.1 by the trace of the ink along the direction of the printing process. However, for conventional offset, patterns 2, 3 are bigger than patterns 4, 5; this is because the contrast of the image is very small, so the trace of the ink almost disappears during the acquisition.

However, with perimeter parameter, it is able to discriminate coated paper and uncoated paper in conventional technology (see Fig. 3.14b). In general, the papers are also distinguishable in waterless process, but not as sufficient as area parameter, (see Fig. 3.14a and Fig. 3.13a). The difference in variance of the perimeter parameter between two substrates is mentioned in sec. 3.2.1.

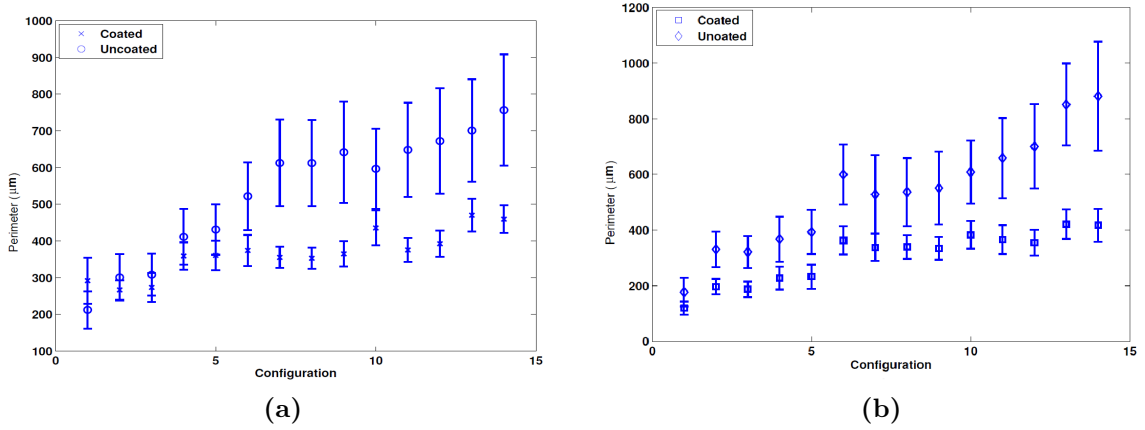


Figure 3.14.: The mean values of the perimeter with waterless process (a) and conventional process (b), resolution 1200 dpi.

Compactness measure of shape, a dimensionless parameter, shown in Fig. 3.15 states that, from an image processing point of view, coated paper gives more compact image than uncoated paper does. This statement is expected, as explained in sec. 3.2.1. In addition, this parameter can be used as a criterion to compare the smoothness/roughness of a surface to another. In the technical view, waterless offset technology, generally, provides less fragmented print than conventional offset does, this is caused by the “wet” process and the characteristic of ink of these two techniques (see sec. 3.1). But these values have so high variance (see Fig. 3.15c,d), so this parameter could not be utilized as a criterion to discriminate the technologies.

Tab. 3.5 represents the behavior of the area ratio in the 1200 dpi resolution. In this case, the values from conventional offset are not negligible any more. In general, the gain in this lower resolution is similar to the higher one. As usual, coated paper shows greater spreading of ink than uncoated paper does. Due to the difference in technology, conventional process with the wet process has reduced the total area of the print.

In the point of view of image analysis, coverage area could be a criterion to discriminate the substrates in waterless offset technology. And again the print from waterless offset is larger than the print from conventional offset. For the conventional technology, the perimeter parameter could be used to differentiate the substrates.

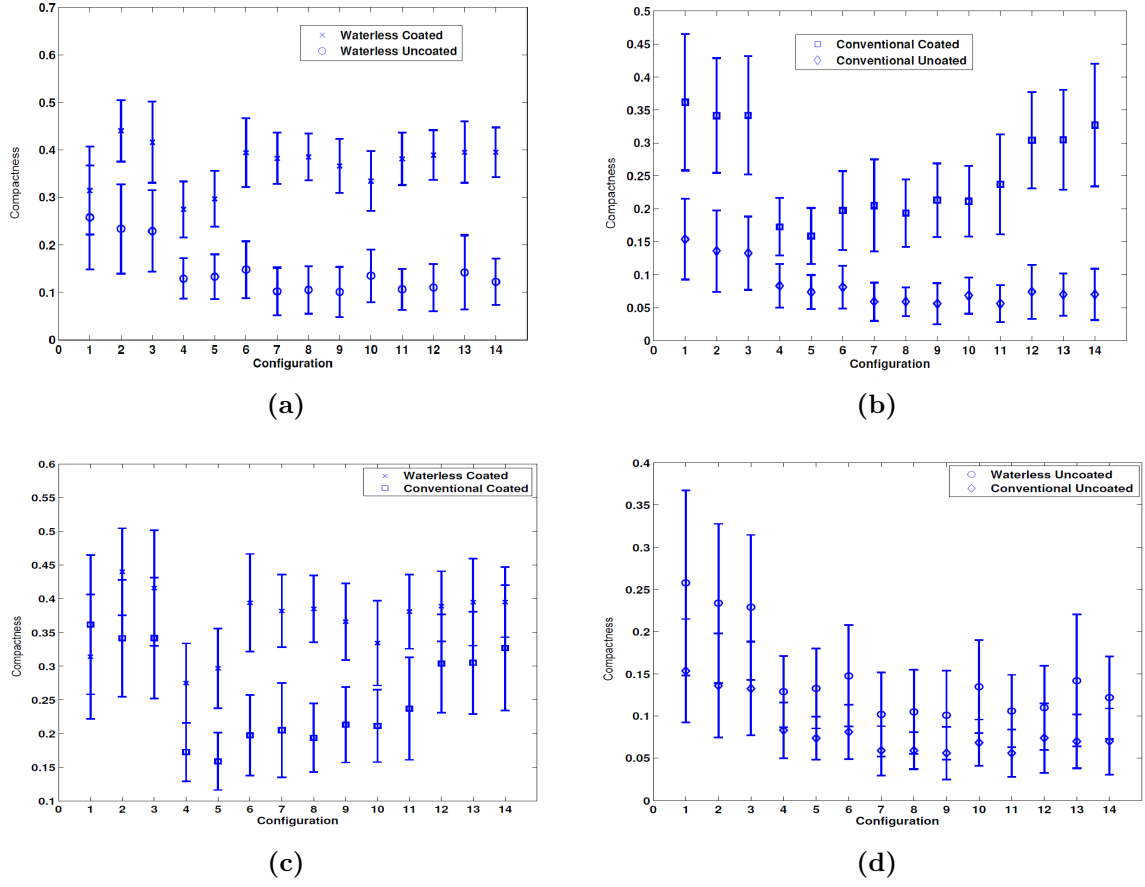


Figure 3.15.: The compactness measure of each pattern with different technologies combine with different substrates, resolution 1200 dpi.

In addition, considering compactness measure of shape, it enables to compare the surface smoothness/roughness of the printing substrates. These parameters are based on the spreading of the ink on the surface and the viscosity of the ink, so that the gain area will provide a lot of information for discrimination. Therefore, the patterns in which the white pixels are covered by many black pixels can make better classification, for instance in this case, patterns 6, 7, 8, 9 or 11.

Tab. 3.9, Tab. 3.10, Tab. 3.11 and Tab. 3.12 give the average images of 100 images of each pattern printed by offset printers on two kind of papers. Focusing on the profile of a single dot, we see that the values are decreasing when its position is going far from the center, this is the fundamental characteristic of the dot shape. Moreover, the images from conventional offset are usually blurrier than the ones from waterless offset; this also happens between two papers as well, the prints on coated paper are generally darker. Between two dots in pattern 5, we can see the trace of the ink between two dots. Looking at pattern 13, the white pixel on the center is totally covered, the print is identical to pattern 14. Waterless offset printer always gives larger coverage than conventional offset printer does.

	Waterless offset		Conventional offset	
	coated	uncoated	coated	uncoated
1	5.63	2.35	1.12	1.04
2	3.48	2.19	1.44	1.54
3	3.40	2.26	1.30	1.44
4	3.89	2.34	0.98	1.20
5	4.25	2.61	0.94	1.25
6	3.04	2.06	1.39	1.53
7	2.69	1.90	1.25	0.83
8	2.68	1.97	1.21	0.89
9	2.71	2.02	1.29	0.86
10	2.79	1.96	1.34	1.06
11	2.40	1.79	1.35	0.98
12	2.23	1.64	1.38	1.18
13	2.42	1.69	1.46	1.29
14	2.07	1.56	1.35	1.16

Table 3.5.: The ratio of average area to theoretical area.

3.2.2. Electrophotography printing

Random structure of paper texture is also an intrinsic “fingerprint”. For instance, laser scanner takes into account the imperfections in the surface to form fingerprint code for the object [7]. In [6], they use the unique non-repeatable features from content to create print signature. Another mean is to collect reflective light in different orientations to construct the 3-D texture [8]. In this experiment with laser printer under the resolution 600 dpi, the focus is on the behavior of the print on two different substrates, coated and uncoated paper. With the same electrophotography printing process, we study the difference of the patterns on two kinds of paper.

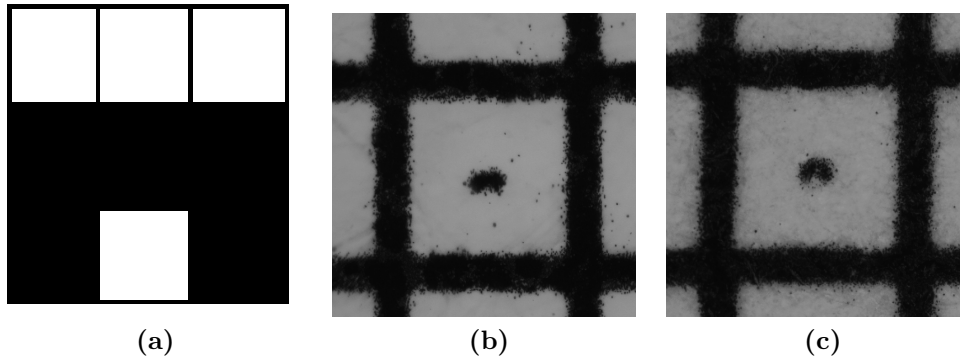


Figure 3.16.: Original print (a), on coated paper (b), on uncoated paper (c), resolution 600 dpi.

The average values of coverage area and perimeter are illustrated in Fig. 3.17. Firstly, considering the area values, the prints on coated paper are greater than the ones on uncoated substrate. However, we observe that the mean values on two

Experimental analysis

substrates are not very separated. Because the ink for laser printer are the toner particles, high viscosity ink, which means that the spreading of these particles is not much on both substrates, the mechanical dot gain is small (see Fig. 3.16). This property is also shown in the results of perimeter values (see Fig. 3.17b) and by the compactness measure parameter (see Fig. 3.18). Nevertheless, when we consider all parameters, the difference between two substrates is more plausible.

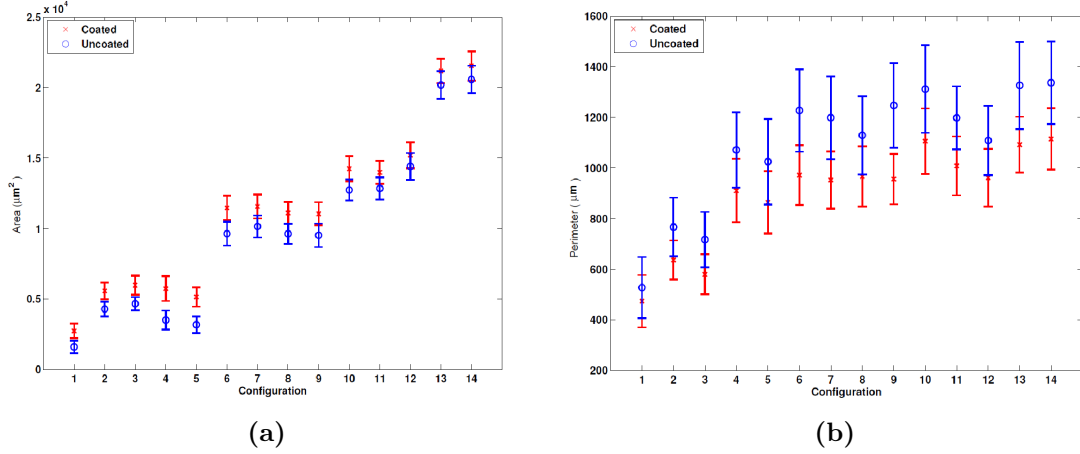


Figure 3.17.: Average values of coverage area (a) and perimeter (b), resolution 600 dpi.

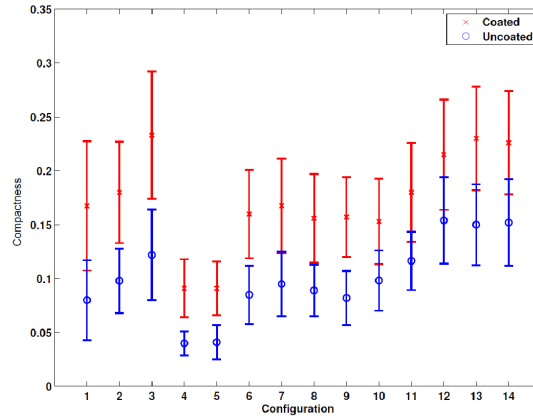


Figure 3.18.: Compactness measure of shape on two different papers, resolution 600 dpi.

Finally, considering the gain area of the print compare to the theoretical parts on the film, with this low resolution of 600 dpi, the gain areas are quite stable (see Tab. 3.6). And, as already mentioned, the covered areas (based on the number of black pixels recognized on the digital images) on the coated paper are mostly greater than the areas gained on rougher-surface paper. They represent twice the ideal area for all patterns.

Tab. 3.7 shows the empirical probabilities of 100 images of single dot, we observe that the emergence frequency of ink at a position depends on the distance to the

Pattern	Coated paper	Uncoated paper
1	1.94	1.13
2	1.98	1.52
3	2.12	1.65
4	2.04	1.24
5	1.82	1.12
6	2.03	1.71
7	2.05	1.80
8	1.97	1.71
9	1.96	1.69
10	2.02	1.81
11	1.98	1.82
12	1.80	1.70
13	2.05	1.79
14	1.96	1.62

Table 3.6.: The ratio of average area to theoretical area, resolution 600 dpi.

center, the further from the center, the lower probabilities of emergence. Moreover, by comparing with Fig. 3.12 and Tab. 3.9, Tab. 3.10, Tab. 3.11, Tab. 3.12, the variation according to the distance is different between the technologies, materials. This frequency of emergence decreases faster with laser technology. The slope of this frequency can characterize the compactness of a dot. In case of many dots, we are concerned by the spreading of the ink when the dots are printed close to each other, the overlap areas are enlarged, Tab. 3.13 and Tab. 3.14. One of the model that takes into account this phenomena is Model-based digital halftoning of Thrasyvoulos N. Pappas [79]. But the model mainly concerns on the optical dot gain which is negligible under the microscope scale, so that it is not appropriate in this case. Therefore, the interaction property is important for the microscopic model.

3.3. Conclusion

From this experimental analysis, the general literature on micro-printed patterns is reviewed. For a particular printing technology and specific material (e.g., ink, substrates,...), the image degradation differs from one to another. The document image degradation models are essential for governing the quality of the printed documents, technically controlling the printing process, the accuracy of document recognition algorithm [35]. The average images calculated from 100 images of one dot from various technologies on coated and uncoated paper are illustrated in Fig. 3.12, Tab. 3.9, Tab. 3.10, Tab. 3.11, Tab. 3.12, Tab. 3.13 and Tab. 3.14. The average profile of an isolated point is rather circular. The experiment gives an overview of the shape of a single dot as well as the shape of other pattern-dots under microscope scale. This observation shows a general difference between printing technologies, the analysis can be used for discriminating the technologies.

This experiment also points out that the distance from the center affects the ink

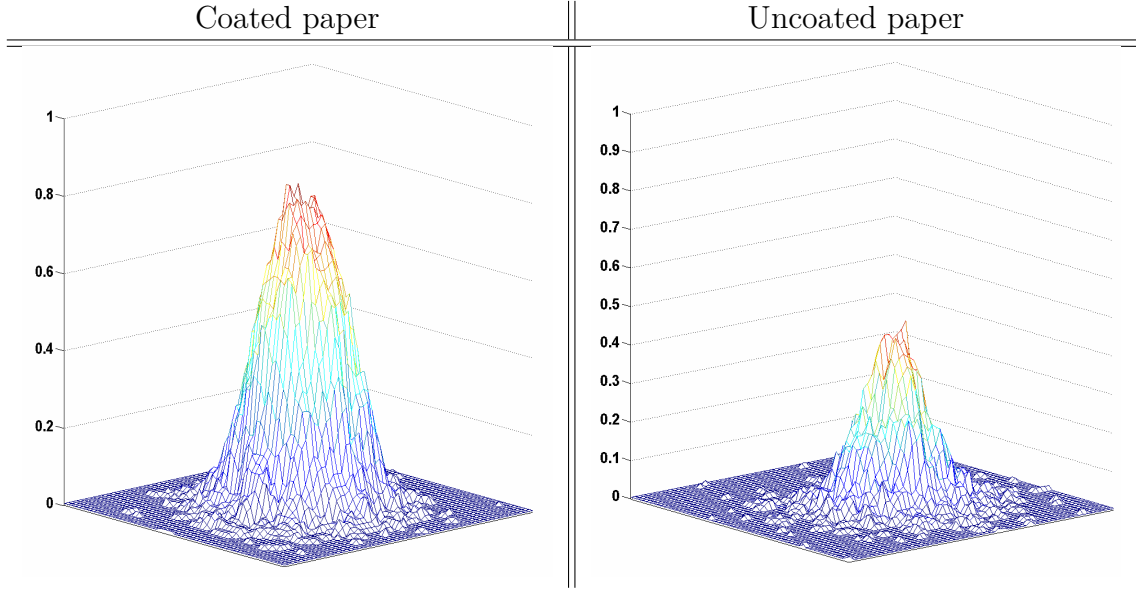


Table 3.7.: The empirical probabilities from 100 samples of one dot from laser printer, resolution 600 dpi.

spreading phenomena. The effect also differs between different processes, this characteristic determines the compactness of the dots. Another characteristic for a print is the interaction between dots which cause the ink spreading on the overlap areas. Based on this observation, in chapter 4, we propose and validate a probabilistic-based model for microscopic prints on the substrate. Studying and understanding the microscopic prints could help us to find the best configuration that allows to discriminate different processes, or/and to find the best configuration to distinguish the correct code from the wrong one.

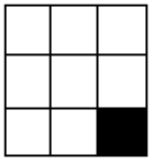
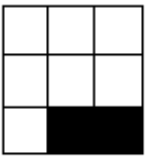
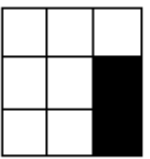
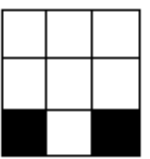
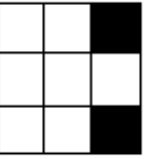
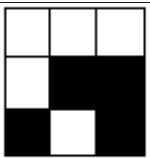
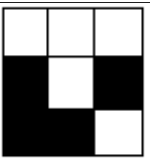
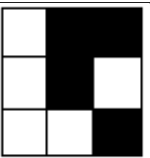
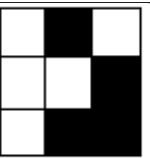

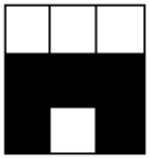
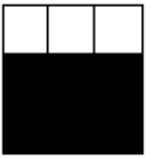


1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.8.: The patterns to observe.

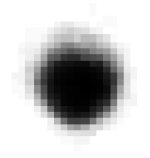
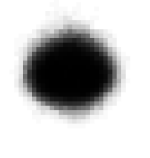
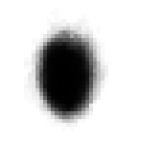
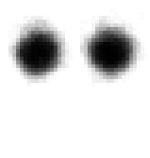
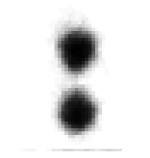
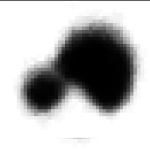
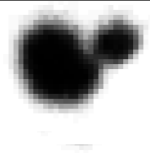
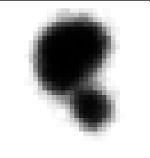
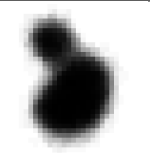
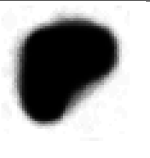
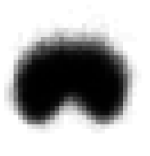
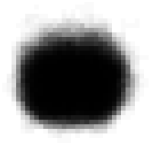
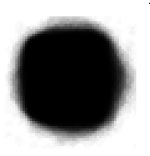
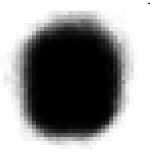
1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.9.: The observed patterns with conventional offset on coated paper, resolution 1200 dpi.

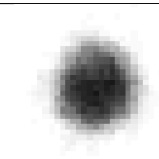
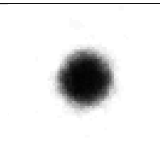
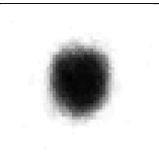
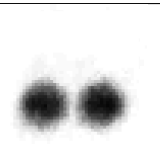
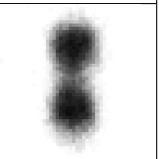
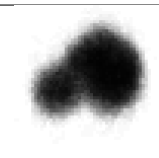
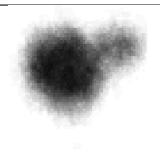
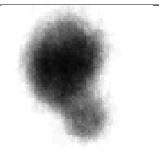
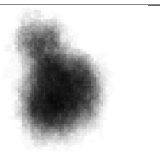
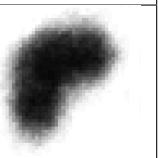
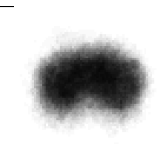
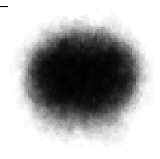
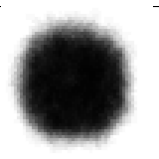
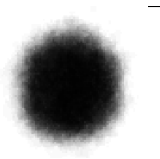
1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.10.: The observed patterns with conventional offset on uncoated paper, resolution 1200 dpi.

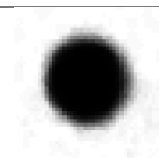
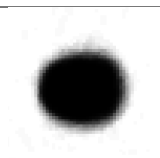
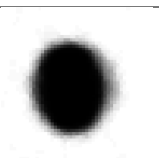
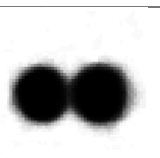
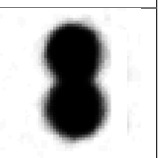
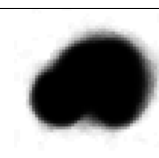
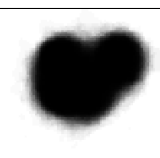
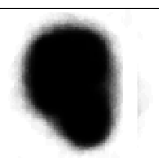
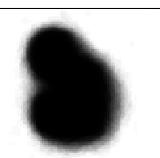
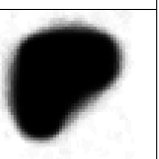
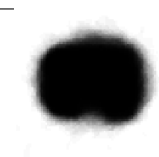
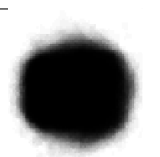
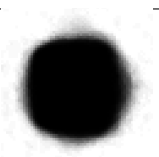
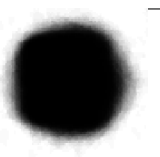
1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.11.: The observed patterns with waterless offset on coated paper, resolution 1200 dpi.

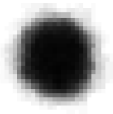
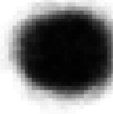
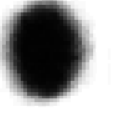




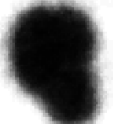
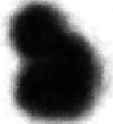


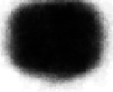

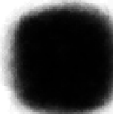
1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.12.: The observed patterns with waterless offset on uncoated paper, resolution 1200 dpi.

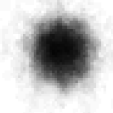
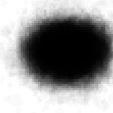
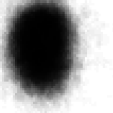









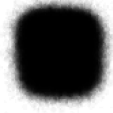

1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.13.: The observed patterns with laser printer on coated paper, resolution 600 dpi.

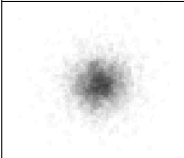
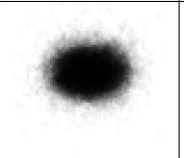
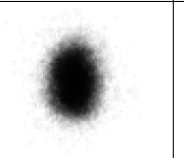
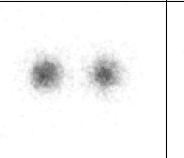
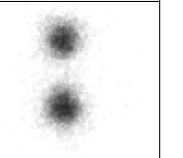
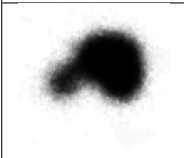
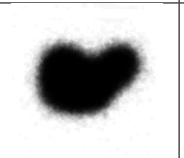
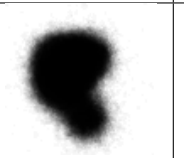
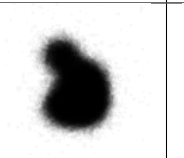
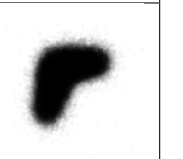
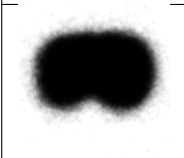
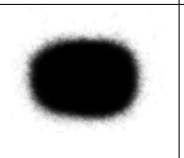
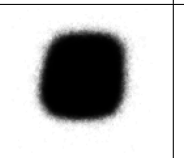
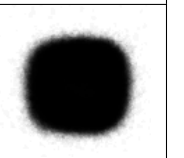
1	2	3	4	5
				
6	7	8	9	10
				
11	12	13	14	
				

Table 3.14.: The observed patterns with laser printer on uncoated paper, resolution 600 dpi.

Chapter 4.

Microscopic printing Probabilistic-based model

From the analysis in chapter 3, the microscopic dots are made of many inked particles potentially unconnected and the classical statistical models of images are inappropriate. In this chapter, we introduce a general model which describes both the gray level distribution and various shapes of dots obtained by various printing technologies. The model includes two main parts: the first one is to construct the parametric distribution of the gray image according to the black and white areas on the print. This gray scale distribution is analyzed, and the Kolmogorov-Smirnov distance is used as a criterion to measure the goodness-of-fit of the parametric distribution. The Pearson system is used to suggest a set of candidates. Second part is to model the hidden binary pixels of the micro print. This binary response model will be investigated in the next chapter.

4.1. Model construction

During the acquisition process, the real image of the printed document is transformed into the approximate digital data, each data unit is called a pixel with an assigned number. If the original document is a binary image, the captured printed-document is stored in the gray-scale level. Let S be a finite set corresponding to the N pixels of the image. The image is modeled by two random processes, $\mathbf{Y} = (Y_s)_{s \in S}$ represents the gray values of the observed image, and $\mathbf{U} = (U_s)_{s \in S}$ constitutes the hidden binary image describing the spatial distribution of the ink (black and white pixels) on the image. Each Y_s takes its value y_s on the positive real line, while the realization u_s of U_s is taken from $\{0, 1\}$, black and white. The parameters of the model are denoted by θ . The model is established as follows

$$P(\mathbf{Y}, \mathbf{U}, \theta) = P(\mathbf{Y} | \mathbf{U}, \theta_{\mathbf{Y}|\mathbf{U}}) P(\mathbf{U} | \theta_{\mathbf{U}}) P(\theta). \quad (4.1)$$

Firstly, $P(\mathbf{Y} | \mathbf{U}, \theta_{\mathbf{Y}|\mathbf{U}})$ models the gray level distribution of the image in the observation vector of parameters $\theta_{\mathbf{Y}|\mathbf{U}}$. This distribution depends on both the substrate and its physical interaction with the ink. The acquisition process also impacts to the final result. Secondly, $P(\mathbf{U} | \theta_{\mathbf{U}})$ models the spatial distribution of the ink at the microscopic scale, a dot is modeled as a set of closed shape depending on the technology of the printer as well as the ink and the paper characteristics. Finally, $P(\theta)$ represents the prior distribution of the parameters $\theta = (\theta_{\mathbf{Y}|\mathbf{U}}, \theta_{\mathbf{U}})$. When θ

is considered as deterministic, it is estimated by a frequentist parameter estimation method such as maximum likelihood estimation, otherwise a Bayesian approach with a cost function is applied.

Particularly, in the sec.4.2, the gray-pixel model $P(\mathbf{Y} | \mathbf{U}, \theta_{\mathbf{Y}|\mathbf{U}})$ is going to be detailed. In the next chapter, the model of the spatial distribution of the ink $P(\mathbf{U} = \mathbf{u} | \theta_{\mathbf{U}})$ is investigated, the estimation methods of parameters $\theta_{\mathbf{U}}$ are also established in this part.

4.2. Ponctual parametric modeling

A captured image from a black and white printed document is a set of gray pixels that take the values in $\{0, 1, 2, \dots, 254, 255\}$. Each pixel value is considered as a random variable that is conditional on its unobserved binary state before printing. We analyze the moments of the data and the histogram shape to propose an approximated parametric model for the histogram.

The ponctual statistic of inked and blank parts is analyzed on two kinds of materials, coated paper and uncoated paper. It is assumed that the random variables $(Y_s)_{s \in S}$ are conditionally independent with respect to $(U_s)_{s \in S}$ and that the distribution of each Y_s is conditional on the state of U_s , e.g.

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{U} = \mathbf{u}, \theta_{\mathbf{Y}|\mathbf{U}}) = \prod_{s \in S} P(Y_s = y_s | U_s = u_s, \theta_{u_s}) := \prod_{s \in S} f_{u_s}(y_s). \quad (4.2)$$

Because u_s is binary, there are only two probability distributions which are f_0 and f_1 . The function f_0 represents the distribution of the gray values of Y_s if the hidden u_s is 0, and f_1 is for the state of white pixel. In the next part, some parametric distributions which are possible for modeling f_0 and f_1 are studied. Many applications dealing with image thresholding are based on the model of f_0 and f_1 , but these approaches mainly used Gaussian distribution [80, 81, 82]. The Log-normal law was also discussed in authentication problem [83, 84], but it is quite intuitive. It is expected to find one or few parametric distributions that approximate more accurately the gray level variation of the printing parts and the non-printing parts on the paper. The statistical analysis consists in the values of the mean, variance, skewness and kurtosis for two kinds of paper, coated paper and uncoated paper.

4.2.1. Theoretical analysis by moments and histogram

Fig.4.3 shows the representative histograms of ink and blank parts. From the observation of the histograms, the distributions that can model the data should have a single peak. Moreover, it indicates that a parametric distribution with shape parameters would be promising, i.e. it can be symmetric or asymmetric according to the values of the shape parameters. The selection of the parametric distributions is based on the shape of the histogram. Particularly, there are two main goals in the section, the first is to propose a method to select a distribution with an adequate shape. The second task is to analyze the moment with regard to various samples. The moment analysis is based on the mean, variance, squared skewness and kurtosis.

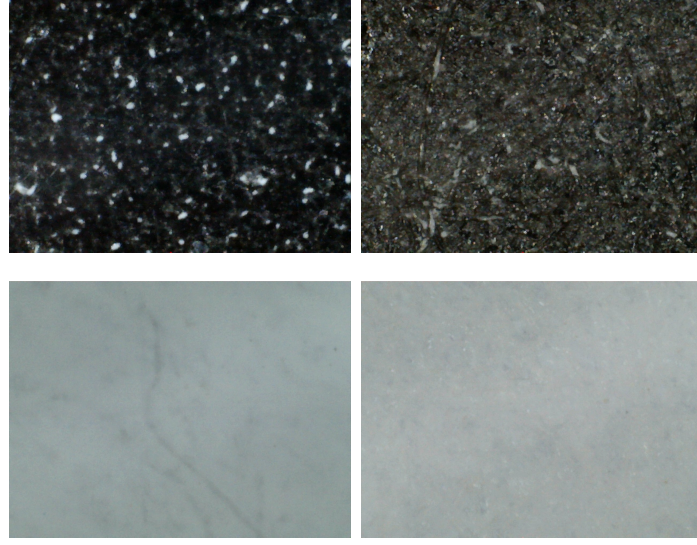


Figure 4.1.: The captured images: coated paper (left), uncoated paper (right); inked area (up), blank area (bottom).

There are many well-known parametric distributions which have been studied. It would be better if one can restrict in only potential candidates. Our approach is based on examining the shape of the histogram with the Pearson system (see sec. A.1) which contains many classes of the popular parametric distributions.

For the set-up of the experiment, the microscope having a magnification of 400 times was used to capture the images. For each kind of sample, 10 images of size 1280×960 have been taken, these samples are printed by conventional offset, waterless offset, laser and inkjet printers on two kind of substrates coated and uncoated papers. These representative samples cover most of the popular printing technologies nowadays. Note that the paper are different for each technology, Tab. 3.1. Tab. 4.1, Tab. 4.2 and Tab. 4.3 show means and variances of the gray values of the samples as well as the squared skewness β_1 and kurtosis β_2 values, estimated by the method of moment, from 10 images of each sample.

Non-printed area	μ	σ^2	β_1	β_2
Coated paper				
Expected value	160.17	12.04	0.04	3.88
Variance	13.49	0.46	0.0009	0.059
Uncoated paper				
Expected value	165.12	13.57	0.53	3.78
Variance	18.74	2.11	0.004	0.39

Table 4.1.: The mean values, variance, squared skewness and kurtosis are calculated from 10 trials of blank area.

For the blank paper, the average gray colors are statistically similar. The skewness values are close to zero, while the kurtosis values are close to three. The histograms have the slightly higher peak and smaller shoulder [85] than the normal

	μ	σ^2	β_1	β_2
Conventional Offset printer				
Expected value	83.04	892.93	0.50	3.32
Variance	39.73	11735	0.076	0.147
Waterless Offset printer				
Expected value	33.64	597.96	4.83	9.86
Variance	26.7	16697	0.983	0.962
Laser printer				
Expected value	38.63	898.5	3.306	5.99
Variance	25.9	42047	0.937	2.03
Inkjet printer				
Expected value	30.56	49.10	2.15	9.50
Variance	1.59	14.09	1.79	24.22

Table 4.2.: The mean values, variance, squared skewness and kurtosis are calculated from 10 trials of printed areas on coated paper.

curve (see Fig. 4.2 and Tab. 4.1), so a close bell-shape distribution is promising to fit the histograms.

According to the Pearson diagram in Fig. 4.2, the indexes of squared skewness of the data of the inked parts are far from zero due to the white speckle noise, see Fig. 4.1. The white speckle is caused by many factors: the non-uniform surface of the substrate, the cohesion within the ink, or the light scattering on the surface. The quality, specially the contrast, of the images is very sensitive to the luminous power.

As seen in Tab. 4.2 and Tab. 4.3, the print on coated paper is generally darker than the print on uncoated paper, this is because the coated layer prevents the ink from being absorbed deeply inside the paper, then makes higher density of ink on

	μ	σ^2	β_1	β_2
Conventional Offset printer				
Expected value	114.48	564.53	0.038	3.25
Variance	27.45	1403.9	0.0018	0.04
Waterless Offset printer				
Expected value	51.28	459.51	2.38	7.44
Variance	8.70	3179.6	0.225	1.18
Laser printer				
Expected value	34.53	493.34	3.02	6.36
Variance	9.63	8296.3	0.337	0.80
Inkjet printer				
Expected value	71.45	435.49	1.15	4.85
Variance	62.05	3335.8	0.2323	0.5

Table 4.3.: The mean values, variance, squared skewness and kurtosis are calculated from 10 trials of printed areas on uncoated paper.

the surface.

Similarly, with the offset printing, the waterless offset process gives darker images than the conventional offset. This result is due to the “wet” process of the conventional offset technology. Also due to humidity, the gray color from the print of conventional printer is more normal and the right tail is lighter than the sample of the other printers, this is why the data from conventional offset printer is more symmetric (Tab. 4.3, Fig. 4.2). With other printers, the histograms have higher peaks and heavier right tails (great values of β_2 [85]), this is due to the aforementioned features that create the white speckles on the print.

On the other hand, the coated paper for inkjet printer is very glossy compared to other coated papers in the experiment. It has a shiny plastic layer that almost prevents the ink from absorbing to the fiber texture. This is the reason why the image is darker and the variance is smaller, see Tab. 4.2.

According to the values of β_1 and β_2 in the Pearson diagram (see Fig. 4.2), the distributions which are analyzed include: Normal, Log-normal, Gamma, Inverse-gamma, Beta distributions. It can be seen in Fig. 4.2 that Inverse-gamma and Log-normal are similar in the presented rank, which means that the goodness-of-fit to the data of these two models are supposed to be close. These are the very popular parametric distributions for a lot of applications. Except the normal law, the other candidates have the skewness and kurtosis indexes controlled by parameters.

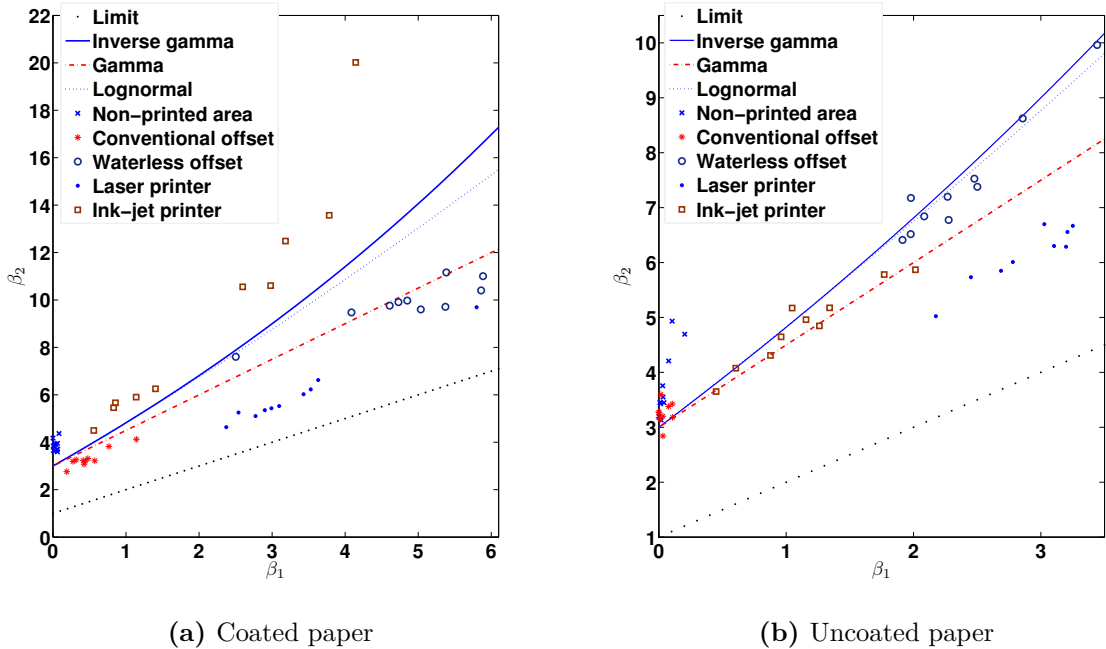


Figure 4.2.: The squared skewness values vs. kurtosis values for the inked and non-inked parts on the diagram of the Pearson system.

4.2.2. Histograms fitting

In this part, the potential parametric distributions are examined to see how do they fit the histograms. Each sample contains 10 captured images of size 1280×960 pixels, then the data used to represent the histogram of a sample has a size of $n = 1280 \times 960 \times 10$ pixels. This is the number of pixels used to estimate the parameters of the candidates for each sample. The applied estimation method is Maximum-likelihood estimation. The Kolmogorov-Smirnov distance (K-S distance) is used as a criterion to measure the absolute difference between the examined distribution and the empirical distribution.

Kolmogorov-Smirnov distance The K-S distance [86] is a simple measure of similarity between two distributions. This measure is defined as the maximum value of the absolute difference between two cumulative distribution functions. The K-S distance between two cumulative distribution functions $F(x)$ and $G(x)$ is defined by

$$\rho_{\infty}(F, G) := \|F - G\|_{\infty} = \sup_x |F(x) - G(x)|. \quad (4.3)$$

The criterion is also used to understand how fit a distribution to a set of observations. The empirical distribution function F_n for n i.i.d. observations $\{x_i\}$ is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1\{x_i < x\}.$$

The Kolmogorov-Smirnov statistic for a given cumulative distribution function $F(x)$ is

$$D_n = \sup_x |F_n(x) - F(x)|. \quad (4.4)$$

If the observations $\{x_i\}$ come from the distribution $F(x)$, the empirical distribution converges to the theoretical one when n goes to infinity.

In the following text, the potential distributions are recalled along with their parameters and properties.

Parametric distributions

Normal distribution Gaussian distribution is interesting because of its numerous applications. It can be used as an approximation to other distributions. As stated by **central limit theorems**, a unit normal distribution is the limit distribution of standardized sums of uncorrelated and identically distributed random variables. The probability density function of normal distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right), \quad (4.5)$$

where $x \in R$, μ is the expectation, σ^2 is the variance. The skewness value of the distribution is zero, $\beta_1 = 0$, which means the pdf is symmetric, and the kurtosis

value β_2 equals to 3. The maximum likelihood estimators of the sample $\{x_i\}$ can be directly obtained. The method of moment estimation also obtains the same estimate formulae. Nevertheless, in order to obtain the unbiased estimators, one usually uses

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i, \\ \hat{\sigma}^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2.\end{aligned}$$

Log-normal distribution A random variable (r.v.) X has log-normal distribution [87] with two parameters σ, μ if the following r.v.

$$U = \frac{1}{\sigma} [\log X - \mu]$$

is a standard normal variable. The probability density function of X is

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} [\log x - \mu]^2 \right\}, \quad x > 0. \quad (4.6)$$

The log-normal distribution models the product of independent and identical positive random variables, $T_n = \prod_{i=1}^n X_i$, in order to achieve a central limit type. Then the limiting distribution of T_n is log-normal. The squared skewness and kurtosis values of the distribution respectively are:

$$\begin{aligned}\beta_1 &= (e^{\sigma^2} + 2)^2 (e^{\sigma^2} - 1) \\ \beta_2 &= e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3.\end{aligned}$$

Unlike the normal distribution, the asymmetry and the peakedness of the log-normal distribution depend on its parameters. It can be seen that, the smaller value of σ^2 the closer the shape of log-normal curve to the normal curve, i.e. β_1 converges to 0 and β_2 converges to 3 when σ^2 goes to zeros. The analytical formulae for the estimators of the parameters using maximum likelihood method are easily obtained by:

$$\begin{aligned}\hat{\mu} &= \frac{1}{n} \sum_{i=1}^n \log x_i, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n \log^2 x_i - \hat{\mu}^2.\end{aligned}$$

Gamma distribution The fit with the Gamma distribution is also highlighted according to the Pearson diagram in Fig.4.2. A random variable X has a gamma distribution (2-parameters) if it has the probability density function of form:

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad (\alpha, \beta > 0, x > 0) \quad (4.7)$$

where α is the shape parameter, β is the scale parameter, and $\Gamma(\cdot)$ is Gamma function. When $\beta = 2$, and $\alpha = k/2$, with k is even, we have the pdf of chi-square distribution with k degrees of freedom. In addition, if U_1, U_2, \dots, U_h are independent unit normal variables, the law of $\sum_{i=1}^h U_i^2$ is also a χ^2 distribution with h degrees of freedom. The squared skewness and kurtosis values of the gamma distribution respectively are:

$$\begin{aligned}\beta_1 &= \frac{4}{\alpha} \\ \beta_2 &= \frac{6}{\alpha} + 3.\end{aligned}$$

Similar to the log-normal distribution, when the shape parameter α is large, the shape of the gamma distribution is similar to the normal curve. In order to achieve maximum-likelihood estimation for α and β , we have to solve numerically the two following equations:

$$\begin{aligned}\beta &= \frac{\bar{x}}{\alpha}, \\ \frac{1}{n} \sum_{i=1}^n \log x_i - \log \bar{x} + \log \alpha - \psi(\alpha) &= 0,\end{aligned}$$

with \bar{x} is the arithmetic mean of $\{x_i\}$.

Inverse-Gamma distribution Fig.4.2 shows that the data are also close to the inverse-gamma distribution, which makes it a potential candidate. The probability density function of inverse-gamma distribution is defined over $x > 0$, [88, 89],

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x}, \quad (\alpha, \beta > 0) \quad (4.8)$$

similar to gamma distribution, α is the shape parameter and β is the scale parameter. If a r.v. X has an inverse-gamma distribution, then its inverse $\frac{1}{X}$ is distributed by a gamma law. The squared skewness and kurtosis values of the distribution respectively are:

$$\begin{aligned}\beta_1 &= \frac{16(\alpha - 2)}{(\alpha - 3)^2} \quad \text{for } \alpha > 3, \\ \beta_2 &= \frac{30\alpha - 66}{(\alpha - 3)(\alpha - 4)} + 3 \quad \text{for } \alpha > 4.\end{aligned}$$

One can observe that the shape of the pdf converges to the shape of a normal curve when α goes to infinity. The maximum likelihood estimators of α and β are obtained by solving numerically the following nonlinear equations:

$$\begin{aligned}\frac{\alpha}{\beta} - \frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} &= 0, \\ \log \alpha - \frac{1}{n} \sum_{i=1}^n \log x_i - \log \left(\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i} \right) - \psi(\alpha) &= 0.\end{aligned}$$

Beta distribution Fig. 4.2 shows some data in type I of Pearson distributions family, which the beta distribution belongs to. The probability density function of beta distribution is defined over $x \in (a, b)$,

$$f(x) = \frac{(x-a)^{\alpha-1}(b-x)^{\beta-1}}{B(\alpha, \beta)(b-a)^{\alpha+\beta-1}}, \quad (\alpha, \beta > 0) \quad (4.9)$$

α and β are the shape parameters, $B(., .)$ is Beta function. The beta distribution was arisen from two independent random variables having gamma distribution, i.e. if X and Y are independent, with $X \sim \Gamma(\alpha, \theta)$ and $Y \sim \Gamma(\beta, \theta)$, then $\frac{X}{X+Y}$ is going to have the standard Beta distribution $Beta(\alpha, \beta)$ with $a = 0, b = 1$. The squared skewness and kurtosis values of the distribution respectively are [90]

$$\begin{aligned} \beta_1 &= \frac{4(\beta - \alpha)^2 (\alpha + \beta + 1)}{(\alpha + \beta + 2)^2 \alpha \beta}, \\ \beta_2 &= \frac{6[(\alpha - \beta)^2 (\alpha + \beta + 1) - \alpha \beta (\alpha + \beta + 2)]}{\alpha \beta (\alpha + \beta + 2)(\alpha + \beta + 3)} + 3. \end{aligned}$$

One can observe that the shape of the pdf converges to the shape of a normal curve when α and β go to infinity simultaneously. Given a and b , the maximum likelihood estimators of α and β are the solutions of the following nonlinear equations which have to be solved numerically:

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \log(x_i - a) - \log(b - a) - [\psi(\alpha) - \psi(\alpha + \beta)] &= 0, \\ \frac{1}{n} \sum_{i=1}^n \log(b - x_i) - \log(b - a) - [\psi(\beta) - \psi(\alpha + \beta)] &= 0. \end{aligned}$$

Experiments

In this part, we verify the fitness of the distributions to the data, and the validation is based on K-S distance criterion. The first goal of the examination is to select one or more good parametric distributions for the modelization. The second goal is to verify how is the dependency of the model on the specific printer. Third target is to answer what is the most appropriate distribution for any kind of print from any printer or technology. Finally, because of the degradation of the printers during time, the consistence of a model for a short period of time should be well concerned. Fig. 4.3 depicts the fit of parametric distributions.

With the ink-free parts, the histograms are almost symmetric (see Tab. 4.1), and all candidates are able to approximate well the data (K-S distances are smaller than 0.06 in all cases). Except for the normal distribution which is symmetric, the others, as mentioned, have a skewness and a kurtosis depending on their parameters. Therefore, the estimators of the parameters will adjust the shape to increase the fitting between the distributions and the histograms.

For instance, with the log-normal distribution, the estimators $\hat{\sigma}^2$ are 0.031 for coated paper and 0.033 for uncoated paper, which makes the distribution closer to the bell-shape. For the gamma curve, because $\hat{\alpha} = 1063.3$ with the coated paper

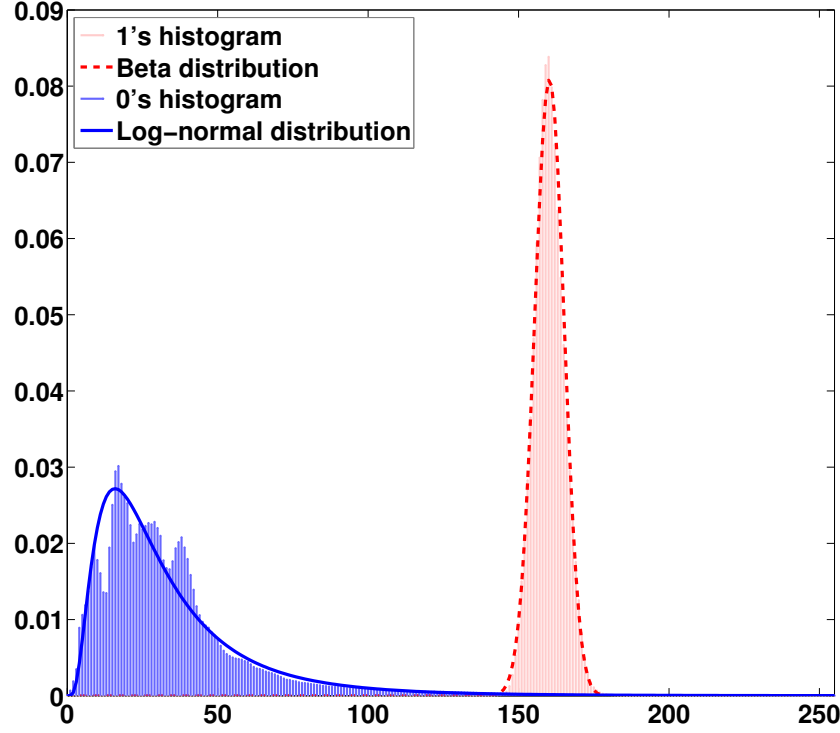


Figure 4.3.: The normalized histograms of black and white areas from waterless offset printer on coated paper, and the fit of distributions.

and $\hat{\alpha} = 884.92$ with the uncoated paper, the parametric distribution is also close to the normal curve. Similarly, the estimation of shape parameter α of inverse-gamma distribution is sufficiently large, $\hat{\alpha}$ are respectively 1065.2 on coated and 873.7 on uncoated. The skewness and kurtosis indexes of inverse-gamma curve are respectively closer to 0 and 3. As expected, the estimators $\hat{\alpha}$ and $\hat{\beta}$ of the Beta distribution are large either ($\hat{\alpha} = 392.5, \hat{\beta} = 232.4$ on coated paper and $\hat{\alpha} = 317.6, \hat{\beta} = 172.9$ on uncoated paper) which makes the shape close to the normal curve. However, in this experiment, the best curve fitting is the **Beta** distribution for both coated and uncoated paper.

With the inked part, the normal curve does not fit well both kinds of paper because it is a symmetric distribution. With log-normal curve, the K-S distances are generally small. Moreover, it is the best distribution for the waterless offset printer, the print from conventional offset printer on coated paper as well as the print on uncoated paper with inkjet printer (see Tab. 4.4b). The Gamma distribution is also a good candidate for the histograms of the inked areas on both papers, (see Tab. 4.4b). The inverse-gamma distribution fits very well the histograms of the prints from laser printer compares to other printers, and the ones on coated paper with inkjet printer. With the beta distribution, in our case, a and b are respectively 0 and 255. Over all, the K-S distances between the beta curves and the histograms are good in all cases. Moreover, it is the best fitting curve for the print from conventional offset

printer on uncoated paper, see Tab. 4.4b.

(a) Non-printing areas

	Coated paper	Uncoated paper
Normal	0.056	0.0417
Log-normal	0.0498	0.0454
Gamma	0.0519	0.0441
Inverse-gamma	0.0478	0.0467
Beta	0.0183	0.0243

(b) Printing areas

	Conventional offset		Waterless Offset	
Paper	Coated	Uncoated	Coated	Uncoated
Normal	0.0723	0.0145	0.1398	0.0923
Log-normal	0.0133	0.0264	0.0316	0.0236
Gamma	0.0236	0.0185	0.0556	0.04
Inverse-gamma	0.0286	0.0362	0.0491	0.0275
Beta	0.0430	0.0143	0.0709	0.0502

	Laser printer		Inkjet printer	
Paper	Coated	Uncoated	Coated	Uncoated
Normal	0.1987	0.1597	0.0966	0.0609
Log-normal	0.0766	0.0494	0.0531	0.0122
Gamma	0.1242	0.0907	0.0681	0.0204
Inverse-gamma	0.0335	0.0188	0.0384	0.0153
Beta	0.1350	0.0952	0.0431	0.0298

Table 4.4.: The Kolmogorov-Smirnov distance between normalized histograms and the tested models.

Tab. 4.5 shows the sum of the K-S distances of all printing technologies and materials for each distribution. It is no surprise that Normal distribution gives the largest value, since it is a symmetric distribution while the histogram is a right-tail curve. The smallest value is given by Inverse-gamma distribution, Log-normal also provides a very good result. The next values respectively are Gamma and then Beta distributions. In general, **Inverse-gamma** law is the most preferable distribution, the second candidate is **Log-normal** distribution which gives the result very close to the inverse-gamma distribution.

Distribution	Normal	Log-normal	Gamma	Inverse-Gamma	Beta
Sum of K-S distances	0.83480	0.2862	0.4411	0.2474	0.4815

Table 4.5.: The sum of K-S distances for each distribution.

However, the choice of a parametric distribution depends also on the specific printer. In particular, Tab. 4.6 shows that the result from HP 600 M602 and Dell laser printers are different from the sample of Xerox and HP color laser printers.

Although they are all laser printers; but the inverse-gamma law has the best fitting curve for the samples from HP 600 M602 and Dell laser printers, while the best choice for the data of HP color and Xerox laser printers is log-normal distribution. This example points out that the different printers can give different results. But the best results are still either from Log-normal distribution or Inverse-gamma law. This conclusion is in line with the previous remark, the shape of these two models are similar in the rank of the data.

	Dell	HP 600 M602	HP color	Xerox
Normal	0.1214	0.1597	0.1088	0.119
Log-normal	0.0347	0.0494	0.0181	0.0144
Gamma	0.0642	0.0907	0.0352	0.0451
Inverse-gamma	0.0143	0.0188	0.0452	0.0184
Beta	0.0914	0.0952	0.0455	0.0703

Table 4.6.: The Kolmogorov-Smirnov distance between normalized histograms and the tested models, samples are taken from four different laser printers.

In addition, Tab. 4.7 illustrates the results of the samples from a frequently used HP laser printer. These samples are printed during three weeks with HP 600 M602 laser printer. These numbers infer that the histograms from the same printer using the same substrate are consistent for a period of time. In particular, this experiment can conclude that the **Inverse-gamma** distribution fits the best the data from the considered HP 600 M602 laser printer. In addition, the order of the quality of each distribution is also unchanged: (2) Log-normal, (3) Gamma, (4) Beta, and (5) Normal distribution.

	1	2	3	4	5
Normal	0.1597	0.1377	0.1007	0.1095	0.0993
Log-normal	0.0494	0.0428	0.0348	0.0346	0.0377
Gamma	0.0907	0.0748	0.0581	0.0606	0.0592
Inverse-gamma	0.0188	0.0201	0.0137	0.0103	0.0177
Beta	0.0952	0.0946	0.0861	0.0886	0.0921

Table 4.7.: The Kolmogorov-Smirnov distance between normalized histograms and the tested models, these 5 samples were collected from the same HP laser printer at different times.

4.3. Conclusion

The general model for the binary gray image has been introduced in (4.1). The model separates the images into two parts, the gray distribution of the blank and printed areas, and the spatial spreading of the inked areas. We can consequently model two parts of the image more efficiently. The model $P(\mathbf{Y} | \mathbf{U}, \theta_{\mathbf{Y}|\mathbf{U}})$ of the gray image has been analyzed. The analysis of the moments of the data with the explanations based on the physical properties has been derived. Moreover, with the

help of the Pearson diagram, a set of selected parametric distributions is proposed to model the data.

Five potential parametric distributions were examined with the K-S distance as a measure of goodness-of-fit, see Tab. 4.4. Since Gaussian law is a symmetric bell-shape distribution, it is difficult to fit the histograms of the inked areas which usually have longer tail on the right. Otherwise, the other distributions have the shape parameters to adjust the skewness and the kurtosis of the curves, which can adapt the right tails of the histograms.

This experiment points out that under a specific condition, more than one distribution can perform well the gray distribution of the print. This performance depends on the specific technology, material, printer that creates the document. With a different condition such as the illumination power, the result could alter.

On the non-printing parts, all candidates can approximate very well the data which have a close bell-shape histogram. And the best candidate belongs to the Beta distribution which achieved the smallest K-S distances to the observations. On the other hand, on the printing parts, aforementioned remark pointed out that there are more than one parametric distribution that fit well observations with a small difference, even from the same technology Tab. 4.6. Nevertheless, it can be concluded that there are two well-fit parametric distributions which are **Inverse-gamma** and **Log-normal** distributions Tab. 4.5. Specifically, Inverse-gamma is the most appropriate parametric distribution that can fit well the gray values histograms of the printed parts in general.

The substrates also make a difference, log-normal distribution is the best proposal for inkjet print on uncoated paper; for coated paper, inverse-gamma distribution give a better fit. Nevertheless, the consistent conclusion can be deduced if the samples are from the same printer and material for a short period of time, e.g. three weeks in our experiment, see Tab. 4.7.

In the authentication point of view, since the results from a particular printer are very consistent, which can be taken advantage by the legitimate producer. The experiment covered most of the popular technologies and materials in the printing community, so that the application is possible [12, 13, 14, 15]. However, the performance of an authentication based on the gray model can be improved if the geometrical properties of the print is also modeled. The analysis of the binary image in a non i.i.d. setting can improve the authentication performance. This is the second part of the model (4.1), which is studied in the next chapters.

Chapter 5.

Spatial binary model

The model of the gray level image \mathbf{Y} in (4.1) has been studied in the previous chapter. We have shown that \mathbf{Y} can be decomposed into gray level distribution of the inked and blanked area, that is $\mathbf{Y} \mid \mathbf{U}$, and a spatial binary process denoted as \mathbf{U} , describing the spatial repartition of the ink on the print. In this chapter, the focus is on the probabilistic modeling of the hidden binary image \mathbf{U} , $P(\mathbf{U} = \mathbf{u} \mid \theta_{\mathbf{U}})$. The model is designed in order to encompass the shapes of the print derived by various printing technologies; e.g. conventional offset, waterless offset, laser or inkjet printers. The proposed model depends on parameters such as the positions of the dots, their spreading and a shape parameter that enables to take into account the variety of printed patterns.

A good modeling requires both a model that fits at best the large variety of observed data and an accurate estimation of the parametric model. In this chapter, we consider the two well-known families of estimation; the frequentist method and the Bayesian one. Because of the complexity of the model, the maximum likelihood method requires an iterative gradient algorithm to reach the maximum. In case of Bayesian method, we resort to a Metropolis within Gibbs algorithm in order to approach the mean squared error criterion. In the results section, the performances of these algorithms are studied from simulated data. Then the modeling and the estimation are assessed from printings derived from the four different technologies.

5.1. Binary response model

In this section, a probabilistic model for the binary image is constructed, and also the properties of the model are provided. Under the microscopic scale, an appropriate model which contains the characteristics of the print is desirable. Let us consider a digital image $\mathbf{U} = (U_s)_{s \in S}$ that displays K dots, each dot is printed independently. In probability theory point of view, since the image is binary, each U_s is considered as a Bernoulli random variable with probability of black p_s . The parameter in the Bernoulli distribution of each pixel s depends on its distances to the centers, denoted by $\{\mu_k\}_{1:K}$. The ability to blacken the surrounding pixels of k -th dot is measured by a function $p_k(\cdot)$. Because a dot is a cluster of ink particles which are aggregated at the center of the dot (Fig. 3.12, Tab. 3.7), $p_k(\cdot)$ decreases continuously with respect to the distance to the center. We consequently propose a function which is inversely proportional to the distance $\|s - \mu_k\|$ and its support is the subset of $(0, 1]$. In the other words, $p_k(s)$ is the parameter of the Bernoulli distribution at

site s . Early assumption is that the random variables $(U_s)_{s \in S}$ are independently distributed, $U_s = 1$ if no ink particle from one of the K dots assigned to the site s , i.e. $U_{s,k} = 1, \forall k \in \{1, \dots, K\}$. On the other hand, the pixel U_s is 0 if at least one $U_{s,k} = 0$. As a consequence, as K dots have an independent effect at site s , U_s is modeled as

$$U_s = \prod_{k=1}^K U_{s,k} \quad (5.1)$$

so that

$$P(U_s = 0) = 1 - \prod_{k=1}^K (1 - p_k(s)), \quad (5.2)$$

where $p_k(s) = P(U_{s,k} = 0)$. Let N be the number of pixels of the image, multivariate random variable $\mathbf{U} = (U_s)_{s \in S}$ is a function $\mathbf{U} : \Omega \rightarrow \{0, 1\}^N$, where Ω is the state space or sample space and $\{0, 1\}^N$ is a N -ary Cartesian power of set $\{0, 1\}$.

Since the random field is composed of independent pixels, the probability mass function $q : \{0, 1\}^N \rightarrow [0, 1]$ for \mathbf{U} is defined as

$$q(\mathbf{u}|\theta) = P(\mathbf{U} = \mathbf{u}) = \prod_s \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right)^{\delta[u_s]} \left(\prod_{k=1}^K (1 - p_k(s)) \right)^{(1-\delta[u_s])} \quad (5.3)$$

where $\delta[\cdot]$ is a function as

$$\delta[u] = \begin{cases} 1, & u = 0 \\ 0, & u \neq 0 \end{cases},$$

and θ is the set of the parameters.

Properties

Number of black pixels Call N_b the number of black pixels of the binary image, N_b is a random variable belonging to $\{0, \dots, \text{card}(S)\}$. Since the process is a set of independent Bernoulli variables of parameter $P(U_s = 0)$, the mean of N_b can be easily deduced

$$n_b = E(N_b) = \sum_{s \in S} E(\delta[U_s]) = \sum_{s \in S} \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right), \quad (5.4)$$

and the variance is

$$\text{Var}(N_b) = \sum_{s \in S} \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) \prod_{k=1}^K (1 - p_k(s)). \quad (5.5)$$

Spread of a single dot Call ρ a quantity measuring the dispersion of the black particles from the center of a single dot. It is defined as

$$\rho = \sqrt{\frac{1}{n_b} \sum_{s \in S} \|s - \mu\|^2 p(s)}, \quad (5.6)$$

where μ is the center of the dot, $p(\cdot)$ is the kernel, and n_b is the average number of black particles of the single dot.

Fisher information The computation of the Fisher Information Matrix is also useful to assess the estimator quality. Under certain conditions, its formula is

$$I(\theta) = -\mathbb{E} \left[\nabla_{\theta} \nabla_{\theta}^T \ln q(\mathbf{U} | \theta) \right]. \quad (5.7)$$

In case of spatial binary model, the Fisher information matrix (5.7) is given by

$$I(\theta) = \sum_{s \in S} \frac{\nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s))\right) \prod_{k=1}^K (1 - p_k(s))}. \quad (5.8)$$

Moreover, due to the relation

$$\nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) = \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta} \ln \left(\prod_{k=1}^K (1 - p_k(s)) \right),$$

the Fisher information matrix is rewritten as

$$I(\theta) = \sum_{s \in S} \frac{\prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s))\right)} \nabla_{\theta} \ln \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \ln \prod_{k=1}^K (1 - p_k(s)) \quad (5.9)$$

One can see that the first factor is the ratio $\Pr(U_s = 1) / \Pr(U_s = 0)$. The detailed calculation of the matrix (5.8) can be found in sec. B.1. Equation (5.9) shows that the Fisher information is the sum of the Fisher information at each pixel. In case of single dot, $K = 1$, the expression is

$$I(\theta) = \sum_{s \in S} \frac{\nabla_{\theta} p(s) \nabla_{\theta}^T p(s)}{p(s) (1 - p(s))}. \quad (5.10)$$

5.2. The Gaussian power kernel

In order to encompass a large variety of shape, a kernel based on the Gaussian function is introduced, namely Gaussian power kernel. We assume here that the kernel $p_k(\cdot)$ has the form

$$p_k(s) = \eta \exp \left[-\frac{1}{2} \left((s - \mu_k)' \Sigma^{-1} (s - \mu_k) \right)^{\beta} \right], \quad (5.11)$$

where η is the height of the peak of the kernel, and μ_k taken in \mathbb{R}^2 is the position

of the center of the peak. At a site s , the Bernoulli parameter $p_k(s)$ depends on its distance from the center of the dot μ_k , $\|s - \mu_k\|_2$. The function decreases when the site s goes further from the center μ_k . Since $p_k(s)$ is a probability, η has to be restricted in the interval $(0, 1]$. The parameter Σ is a positive definite symmetric matrix, it controls the largeness of the droplets and the ellipse shape of the level sets of the Gaussian. The parameter β controlling the shape of $p_k(\cdot)$ is a positive real number, particularly, it controls the density of the black particles. It is easy to see that with $\beta = 1$, (5.11) becomes the classical Gaussian function. In our problem, shape parameter β plays an important role to characterize an appropriate shape for the micro printed dot. Fig. 5.1 shows the shape of $p_k(s)$ in different values of β . The vector of parameters is then $\theta_U = [\{\mu_k\}, \Sigma, \beta, \eta]$.

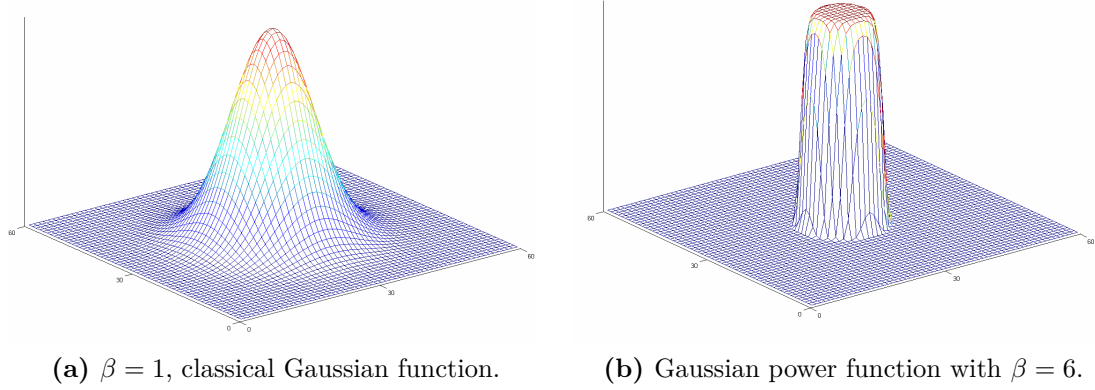


Figure 5.1.: Gaussian power function with varied β in \mathbb{R}^2 , $K = 1$.

In addition, since (5.4) gives the expected number of black pixels, the formula with a single dot, $K = 1$, is:

$$n = \sum_s p_1(s). \quad (5.12)$$

Then a parameter n controlling the quantity of black pixels of one dot is unveiled. This parameter represents the expected number of black pixels that can be generated by a single dot. Assuming that the kernel $p_k(\cdot)$ of k -th dot has the form

$$p_k(s) = n f(s | \theta_k), \quad (5.13)$$

while f has the properties of a probability density function. As a result, the expected number of black pixels of the k -th dot is

$$\sum_s p_k(s) = n \sum_s f(s | \theta_k) \approx n.$$

Then the parameter θ_k of f contributes mainly to the formation of the shape of p_k . From (5.13) and (5.11), η is factorized into the expected number of black pixels n

of a dot and a normalizing factor. The form of the density function f is

$$f(s \mid \mu, \Sigma, \beta) = \frac{1}{|\Sigma|^{\frac{1}{2}} \pi \Gamma\left(1 + \frac{1}{\beta}\right) 2^{\frac{1}{\beta}}} \exp \left\{ -\frac{1}{2} \left((s - \mu)' \Sigma^{-1} (s - \mu) \right)^{\beta} \right\}. \quad (5.14)$$

This is the density function of the exponential power (EP) distribution [91, 92]. It is noted that f is a multivariate normal density function since $\beta = 1$. As $\beta \rightarrow \infty$, the density converges pointwise to a uniform density. It is considered as an extensions of the normal distribution and helpful in modeling random phenomena [91]. Due to the more general version of Gaussian distribution, it can have wide applicability, and can approximate the data more precisely. However, when the shape parameter β is included, the estimation has to handle the gamma function which is an improper integral. This issue makes the estimation more complicated, then in many applications, when the difference is not significant, the Gaussian distribution is used instead. On the other hand, as mentioned, shape parameter β defines an appropriate shape for the micro printed dot in our problem.

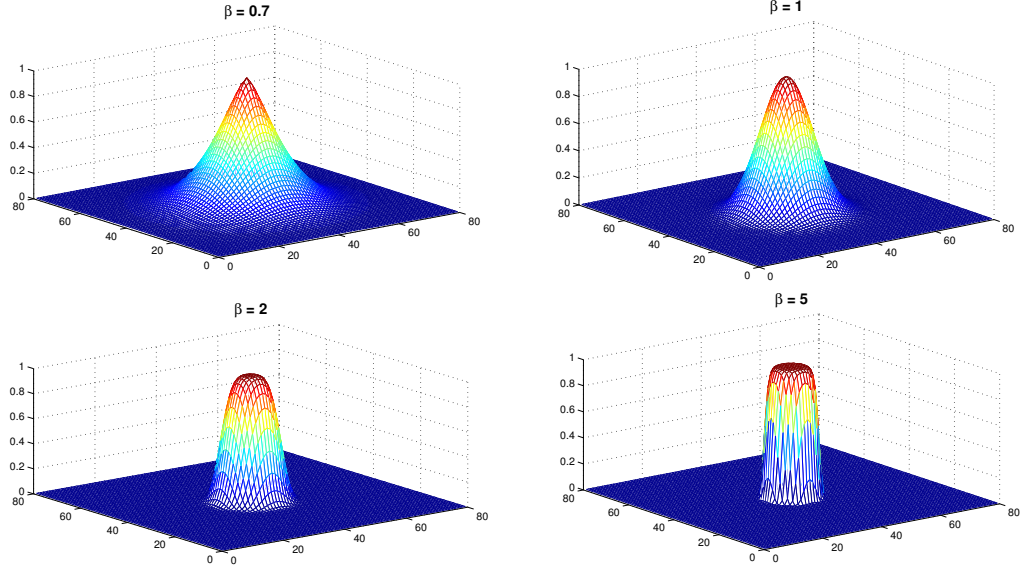
Σ , β and n depend on the printer, its technology and its tuning. n in (5.13) plays the role as the tone of ink, it controls the amplitude of $p_k(s)$. Practically, the Gaussian power kernel $p_k(s)$ is assumed to be circular, i.e. $\Sigma = \sigma^2 I_2$ with I_2 the 2-dimensional identity matrix

$$p_k(s \mid \mu_k, \sigma, \beta, n) = \frac{n}{\pi \sigma^2 2^{\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right)} \exp \left\{ -\frac{1}{2} \left(\frac{\|s - \mu_k\|_2^2}{\sigma^2} \right)^{\beta} \right\}. \quad (5.15)$$

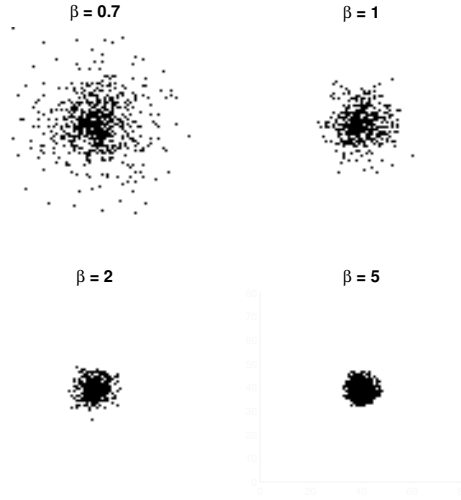
The expected number of particles n is restricted in $(0, \pi \sigma^2 2^{\frac{1}{\beta}} \Gamma(1 + \frac{1}{\beta})]$ because η belongs to $(0, 1]$. The spread of a single dot in this case is then $\rho = \sqrt{\sum_{s \in S} \|s - \mu\|^2 f(s)}$, from the formula of the variance of the exponential power random variable (Proposition 3.2 in [91]), we obtain

$$\rho \approx \frac{\sigma^2 2^{\frac{1}{\beta}} \Gamma\left(\frac{2}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)}. \quad (5.16)$$

The parameters of the binary response model are denoted $\theta_U = \{\mu_1, \dots, \mu_K, \sigma^2, \beta, n\}$. Fig. 5.2 shows both $p_k(s)$ and dots in case of four different values of β .



(a) $p_k(\cdot)$ with $\sigma^2 = 50$, $n = 2^{\frac{1}{\beta}} \pi \sigma^2 \Gamma\left(1 + \frac{1}{\beta}\right)$ and $\beta \in \{0.7; 1; 2; 5\}$



(b) realizations of the binary response model

Figure 5.2.: The shape of kernels and their realization.

Consider the model of single dot, i.e. $K = 1$, model (5.3) is rewritten as

$$q(\mathbf{u}|\theta) = \prod_s (1 - p_1(s))^{\mathbb{I}_{S_0}(s)} (1 - p_1(s))^{\mathbb{I}_{S_1}(s)}. \quad (5.17)$$

Using (5.10), we calculate the Fisher information matrix $I(\theta)$ to compute the Cramér-

Rao lower bounds. The elements on the diagonal of the matrix is calculated as

$$\begin{aligned}
I_{\mu_i, \mu_i} &= \frac{\beta^2}{\sigma^{4\beta}} \sum_{s \in S} r_s (s_i - \mu_i)^2 \|s - \mu\|^{4\beta-4}, \\
I_{\sigma^2, \sigma^2} &= \frac{1}{\sigma^4} \sum_{s \in S} r_s \left[\frac{\beta}{2} \left(\frac{\|s - \mu\|}{\beta} \right)^\beta - 1 \right]^2, \\
I_{\beta, \beta} &= \sum_{s \in S} r_s \left[\frac{\ln 2 + \psi\left(1 + \frac{1}{\beta}\right)}{\beta^2} - \frac{1}{2} \left(\frac{\|s - \mu\|^2}{\sigma^2} \right)^\beta \ln \left(\frac{\|s - \mu\|^2}{\sigma^2} \right) \right]^2, \quad (5.18) \\
I_{n, n} &= \frac{1}{n^2} \sum_{s \in S} r_s,
\end{aligned}$$

with $i = 1, 2$ and $r_s = \frac{p_1(s)}{1-p_1(s)}$. The detail of the calculation and other elements of the matrix can be found in sec. B.2. From (5.13), one can see that r_s increases with n , on the other side, the Fisher information matrix can be expressed as $I = \sum_{s \in S} r_s M_s$. This means that the more number of black pixels we have, the more information for the estimation procedure we are able to get.

The Cramér-Rao lower bounds of the variances of the unbiased estimators are shown in Fig. 5.5 and Fig. 5.6. These figures show that it is possible to achieve smaller variance of estimators of μ, σ^2 and n with larger β . When β is close to 0, the lower bounds are very large, in the other words, the estimation for μ, σ^2 and n is very difficult. In contrast, the lower bounds of the variance of the estimator $\hat{\beta}$ gets larger when true β is large, see Fig. 5.6a. This phenomena is reasonable, because the density functions (5.14) are less distinguishable with the large values of β . This means that the realizations with large β look likely the same.

The Kullback-Leibler divergence is used as an asymmetric similarity measure, named KL distance, between density functions. The distribution $q(u | \theta_U)$ is defined based on the kernels p_k . Under the assumption (5.13), we can see that with the same n , the more difference between the probability density functions f , the more difference between the distributions q . Let us define f_β as a distribution whose density function is of form $f(\cdot | \mu, \sigma, \beta)$, with fixed μ, σ , then the KL distance between f_{β_0} and f_β is obtained as following [93]

$$\begin{aligned}
D_{KL}(f_\beta \| f_{\beta_0}) &= E_{f_\beta} \left[\frac{\ln f_\beta(\mathbf{S})}{\ln f_{\beta_0}(\mathbf{S})} \right] \\
&= \ln \frac{\beta \Gamma\left(\frac{1}{\beta_0}\right)}{\beta_0 \Gamma\left(\frac{1}{\beta}\right)} + \left(\frac{1}{\beta_0} - \frac{1}{\beta} \right) \ln 2 + \frac{2^{\frac{\beta_0}{\beta}-1} \Gamma\left(\frac{\beta_0+1}{\beta}\right)}{\Gamma\left(\frac{1}{\beta}\right)} - \frac{1}{\beta}. \quad (5.19)
\end{aligned}$$

Set $\beta = \beta_0 + \delta$, with fixed δ , then the KL distance goes to zero when β_0 goes to infinity, the calculation of the limit is presented in sec. B.3. Fig. 5.4 illustrates the KL distance between f_β and f_{β_0} , the curves respect to three values of δ are shown. The figure indicates that the distributions are more unlikely to differentiate for great β . Since the distribution $q(u | \theta_U)$ depends on p_k whose shape is defined by f_β , the

difference between distributions q are also small with large β . Fig.5.3 shows the realization with three values of β .

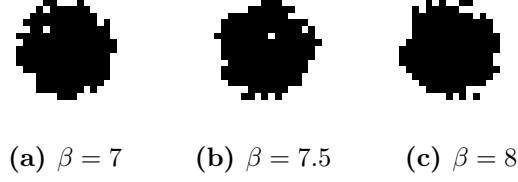


Figure 5.3.: Realization with three different values of β .

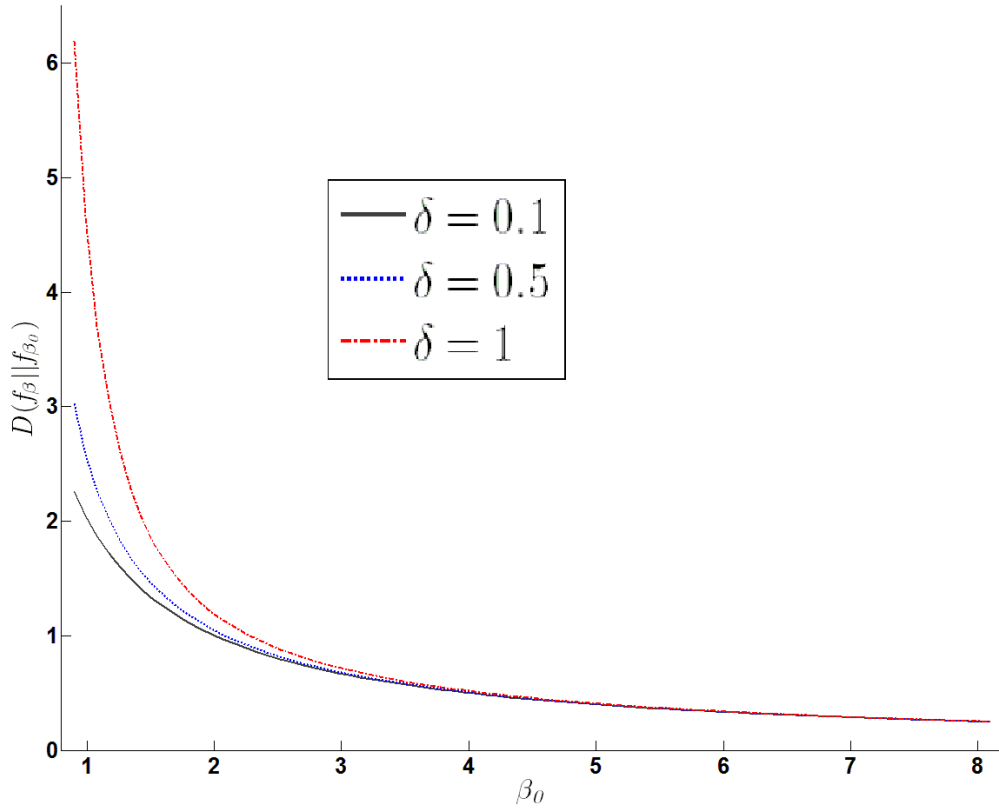


Figure 5.4.: KL distance between density functions with different β , $\beta = \beta_0 + \delta$.

5.3. Estimation methods

5.3.1. Maximum Likelihood estimation

In this part, the maximum likelihood estimation is described. Since \mathbf{U} is made of independent random variables, the log-likelihood of all pixels is given by

$$L(\theta) = \sum_{s \in S_0} \ln \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) + \sum_{s \in S_1} \sum_{k=1}^K \ln (1 - p_k(s)), \quad (5.20)$$

with $S_i = \{s \mid u_s = i\}$. With the kernel (5.11), the parameters σ^2, β and η have to be consistent with:

$$-\sigma^2 \leq 0 \quad (5.21)$$

$$-\beta \leq 0 \quad (5.22)$$

$$\eta - 1 \leq 0 \quad (5.23)$$

$$-\eta \leq 0 \quad (5.24)$$

The right sides of (5.21), (5.22), (5.23), (5.24) are respectively denoted by $g_1(\theta), g_2(\theta), g_3(\theta), g_4(\theta)$. Because the ML estimator for the model parameters is not analytically tractable, an optimization approach with an iterative algorithm consisting in successive estimation of each parameters is derived. The initialization of the algorithm is a first rough approximation of θ . The initialization of the means is obtained as follow

- The first initial center μ_1^g is chosen uniformly at random from the set of black pixels, S_0
- Compute $D(s)$, the distance from a black pixel s to the closest chosen center(s)
- Take a next center μ_k^g , choose s from S_0 black pixels with probability $\frac{D(s)^2}{\sum_{s \in S_0} D(s)^2}$
- Proceed until K initial centers are collected.

This careful seeding helps to reduce the chance that the centers are initialized so close to each others, the method was proposed in [94].

Denote N_0 the number of the black pixels of the image. Since σ controls the width of the dot, similar to the radius of a circle with N_0 is considered as the area. With this relation, it is guessed as

$$\sigma^{2g} = \frac{N_0}{K\pi}. \quad (5.25)$$

The other parameter β and η are initialized with respect to a specific case.

At each iteration of the algorithm, the means are estimated from the quasi-Newton algorithm (QNA) based on the BFGS method, sec. 2.3.1. While β, σ and η have the constraints (5.21), (5.22), (5.23), (5.24), the augmented Lagrangian method with the QNA is applied, sec. 2.3.1. At each iteration i , the parameters are estimated separately in turn, this procedure is called alternating algorithm (see chapter 10 in [95] or [96, 97]). In particular, $\{\mu_k^{(i)}\}$ are the argument of the maximum of the function $L(\cdot)$ with other parameters fixed from iteration $i - 1$,

$$\{\mu_k^{(i)}\} = \arg \max L(\mu_1, \dots, \mu_K \mid \sigma^{2(i-1)}, \beta^{(i-1)}, \eta^{(i-1)}).$$

Then the estimator $\sigma^{2(i)}$ is estimated using the recent $\{\mu_k^{(i)}\}$ and $\beta^{(i-1)}, \eta^{(i-1)}$ with constraint (5.21),

$$\sigma^{2(i)} = \arg \max L(\sigma^2 \mid \{\mu_k^{(i)}\}, \beta^{(i-1)}, \eta^{(i-1)}).$$

The procedure is similar for the other parameters,

$$\begin{aligned}\beta^{(i)} &= \arg \max L\left(\beta \mid \left\{\mu_k^{(i)}\right\}, \sigma^{2(i)}, \eta^{(i-1)}\right) \quad \text{with constraints (5.22)} \\ \eta^{(i)} &= \arg \max L\left(\eta \mid \left\{\mu_k^{(i)}\right\}, \sigma^{2(i)}, \beta^{(i)}\right) \quad \text{with constraints (5.23), (5.24)}.\end{aligned}$$

Then estimation of σ^2 consists in minimizing the function $C(\sigma^2, \lambda, r)$

$$C(\sigma^2, \lambda, r) = -L(\sigma^2) + \lambda g_1(\sigma^2) + \frac{r}{2} \max(0, g_1(\sigma^2))^2, \quad (5.26)$$

where $\lambda \geq 0$ and $r \geq 0$ are respectively the Lagrangian multiplier and the penalty parameter. The initialization of this step is such that $r^{(0)} > 1$ and $\lambda^{(0)} \geq 0$. The iteration j of the minimization algorithm proceeds as follows:

- **New approximation:** Minimizing the augmented Lagrangian $C(\sigma^2, \lambda^{(j)}, r^{(j)})$ with respect to $\sigma^{2(j)}$ from the QNA with BFGS method
- **Stopping criterion:**

$$\begin{aligned}\left| -\partial_{\sigma^2} L(\sigma^{2(j)}) + \lambda^{(j)} \partial_{\sigma^2} g_1(\sigma^{2(j)}) \right| &< \epsilon, \\ \left| \sigma^{2(j)} g_1(\sigma^{2(j)}) \right| &< \epsilon\end{aligned}$$

- **Update:**

$$\begin{aligned}\lambda^{(j+1)} &= \max\{0, \lambda^{(j)} + r^{(j)} g_1(\sigma^{2(j)})\} \\ r^{(j+1)} &= cr^{(j)}\end{aligned}$$

where c is usually chosen large to boost the penalty functions. At each iteration i , the estimation of the other parameters follows the same method, only the augmented Lagrangian function C is modified to take into account the specific constraint(s) of each parameter.

In addition, when the exponential power kernel (5.15) is taken into account, the parameters in the log-likelihood function (5.20) σ^2, β and n now have to satisfy:

$$n - 2^{\frac{1}{\beta}} \pi \sigma^2 \Gamma\left(1 + \frac{1}{\beta}\right) \leq 0 \quad (5.27)$$

$$-n \leq 0 \quad (5.28)$$

$$-\beta \leq 0. \quad (5.29)$$

The initialization will be modified slightly. Particularly, the initialization of the means is unchanged. However, since n stands for the average number of black particles of a dot, it is natural to initialize it from the number of back pixels N_0 , i.e.

$$n^g = \frac{N_0}{K}. \quad (5.30)$$

Practically, σ^2 is initialized by setting the constraint (5.27) as equal. The term

$\left(2^{\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right)\right)^{-1}$ is in $(0, 1)$, with experience, we see that it should be close to 1,

$$\sigma^{2g} = 0.8 \times \frac{n^g}{\pi}. \quad (5.31)$$

Also from (5.27), $\Gamma\left(1 + \frac{1}{\beta}\right)$ is close to 1 when $\beta > 1$, then

$$\beta^g = \frac{\ln 2}{\ln n^g - \ln(\pi \sigma^{2g})}. \quad (5.32)$$

There is a remark here, the initialization in case of exponential power kernel can be taken advantage to initialize η and β in case of general Gaussian power kernel more efficiently. Specifically, the initial of β will be (5.32), while η can be initialized by the fact that η is factorized into n divided by a normalizing factor (see (5.15) and (5.11)). The clusters blind identification algorithm with Gaussian power kernel is summarized in Algorithm 5.1.

Algorithm 5.1 Gaussian power binary response model identification.

Initialize $\theta^{(0)} = [\mu_1^{(0)}, \mu_2^{(0)}, \dots, \mu_K^{(0)}, \sigma^{2(0)}, \beta^{(0)}, \eta^{(0)}]$

Choose ϵ

while $|L(\theta^{(j-1)}) - L(\theta^{(j)})| > \epsilon$ **do**

Estimate $\{\mu_k^{(j)}\}$ by BFGS method with $(\sigma^{2(j-1)}, \beta^{(j-1)}, \eta^{(j-1)})$

Estimate $\sigma^{2(j)}$ by augmented Lagrangian method with $(\{\mu_k^{(j)}\}, \beta^{(j-1)}, \eta^{(j-1)})$ and constraint (5.21)

Estimate $\beta^{(j)}$ by augmented Lagrangian method with $(\{\mu_k^{(j)}\}, \sigma^{2(j)}, \eta^{(j-1)})$ and constraints (5.24)

Estimate $\eta^{(j)}$ by augmented Lagrangian method with $(\{\mu_k^{(j)}\}, \sigma^{2(j)}, \beta^{(j)})$ and constraints (5.22), (5.23)

end while

5.3.2. Bayesian estimation method

In this case, the estimation method of the parameters based on Bayesian inference is derived. Let us recall the problem, the interesting parameters are $\theta_{\mathbf{U}} = [\{\mu_k\}, \sigma^2, \beta, \eta]$ which are considered as the random variables instead of deterministic values. The method of Minimum Mean Squared Error is applied, the estimator $\hat{\theta}$ of $\theta_{\mathbf{U}}$ given the observed data \mathbf{u} is

$$\hat{\theta}(\mathbf{u}) = \int \theta p(\theta | \mathbf{u}) d\theta \quad (5.33)$$

where $p(\theta | \mathbf{u})$ is the posterior distribution. To approximate the integral (5.33), the MCMC algorithm is applied. Since the vector of parameters is of high-dimension, it is more convenient to use Gibbs sampling algorithm. In particular, the estimation is based on Metropolis-Hastings within Gibbs algorithm described by Algorithm 2.5.

Moreover, the interested parameters are considered to be independent. The chosen strategy is to run a random walk (2.24) to generate the Markov chain. Particularly, the normal distribution is used for the random walk, such that the proposal function is symmetric. Since σ^2, β are positive, we generate $\theta_2 = \log(\sigma^2)$ and $\theta_3 = \log(\beta)$ instead of the parameters itself. For η , it is between 0 and 1, it would be wise to apply the following homeomorphism from \mathbb{R} to $(0, 1)$

$$\eta = \frac{1}{1 + e^x}. \quad (5.34)$$

For convenient notation, the vector of interested parameters are rewritten as $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$, where θ_1 stands for $\{\mu_k\}$, and θ_4 is x in (5.34). The Metropolis-Hastings within Gibbs using random walk is summarized in Algorithm 5.2.

Algorithm 5.2 Sampling algorithm with random walk.

Initialize $\theta^{(0)} = [\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}]$

for $i = 1, \dots, N_iter$ **do**

for $b = 1, \dots, 4$ **do**

 Generate $\theta_b^* \sim \mathcal{N}(\theta_b^{(i-1)}, \sigma_b^2)$

 Compute the acceptance ratio

$$\alpha(\theta_b^{(i-1)}, \theta_b^*) = \min \left\{ \frac{p(\theta_b^* | \theta_{-b}^{(i-1)}, \mathbf{u})}{p(\theta_b^{(i-1)} | \theta_{-b}^{(i-1)}, \mathbf{u})}, 1 \right\}$$

 Decide

$$\theta_b^{(i)} = \begin{cases} \theta_b^* & \text{with probability } \alpha(\theta_b^{(i-1)}, \theta_b^*) \\ \theta_b^{(i-1)} & \text{otherwise} \end{cases}$$

end for

end for

Recall that $\theta_{-b}^{(i-1)} = [\theta_1^{(i-1)}, \dots, \theta_{b-1}^{(i-1)}, \theta_{b+1}^{(i-1)}, \dots, \theta_4^{(i-1)}]$. The initialization θ is obtained by the careful seeding for the means. The other parameters are initiated similarly to MLE method. The prior distributions are chosen with fixed parameters by experience. Each center $\{\mu_k\}$ has **multivariate normal** distribution. Since σ^2 has to be positive, it is assumed to have a **log-normal distribution**, β has **gamma distribution**, while η is **beta-distributed**. From the Bayes's law, we have

$$p(\theta_b | \theta_{-b}, \mathbf{u}) = \frac{p(\mathbf{u} | \theta_b, \theta_{-b}) p(\theta_b, \theta_{-b})}{p(\theta_{-b}, \mathbf{u})},$$

then

$$\begin{aligned} \frac{p(\theta_b^* | \theta_{-b}^{(i-1)}, \mathbf{u})}{p(\theta_b^{(i-1)} | \theta_{-b}^{(i-1)}, \mathbf{u})} &= \frac{p(\mathbf{u} | \theta_b^*, \theta_{-b}^{(i-1)}) p(\theta_b^*, \theta_{-b}^{(i-1)})}{p(\theta_{-b}^{(i-1)}, \mathbf{u})} \frac{p(\theta_{-b}^{(i-1)}, \mathbf{u})}{p(\mathbf{u} | \theta_b^{(i-1)}, \theta_{-b}^{(i-1)}) p(\theta_b^{(i-1)}, \theta_{-b}^{(i-1)})} \\ &= \frac{p(\mathbf{u} | \theta_b^*, \theta_{-b}^{(i-1)}) p(\theta_b^*, \theta_{-b}^{(i-1)})}{p(\mathbf{u} | \theta_b^{(i-1)}, \theta_{-b}^{(i-1)}) p(\theta_b^{(i-1)}, \theta_{-b}^{(i-1)})}, \end{aligned}$$

where $p(\mathbf{u} | \theta_b, \theta_{-b})$ is the likelihood and $p(\theta_b, \theta_{-b})$ is the prior. Since each parameter is distributed independently, the acceptance ratio in Algorithm 5.2 is rewritten as

$$\alpha(\theta_b^{(i-1)}, \theta_b^*) = \min \left\{ \frac{p(\mathbf{u} | \theta_b^*, \theta_{-b}) p_b(\theta_b^*)}{p(\mathbf{u} | \theta_b^{(i-1)}, \theta_{-b}) p_b(\theta_b^{(i-1)})}, 1 \right\}, \quad (5.35)$$

with p_b the prior of parameter θ_b . Moreover, when the other parameters are known, there is always a formula to compute n from η and vice versa (5.15). So that the Bayesian estimation for the case of exponential power kernel is not necessary. Let us note that when the priors are uniform or when the variance of the priors are large, the ratio of priors in (5.35) is close to one and the acceptance ratio mainly depends on the likelihood ratio.

5.4. Experimental results and performance

In this section, we aim at assessing both the performance of the estimators and the adequacy of the model. The maximum likelihood estimators and the Bayesian ones are analyzed, their performance are evaluated from the images generated by the parametric model. Tab. 5.1 visualizes the database of the simulated dots with different values of the parameters. It is used to evaluate the performance of the algorithms.

Then we study the adequacy of the estimated model with few printings coming from various printers and technologies. The parametric model is justified by the images obtained from four printing technologies by two estimation methods. We point out the strength and the weakness of the model for modeling printing dots of printers.

5.4.1. Maximum likelihood estimation

The performance of the Maximum Likelihood estimation is analyzed at first with simulated data. In particular, in this experiment, the impact of the various values of σ^2, β, n on the performance of the estimators is investigated. Their variances are compared to the Cramér-Rao bound. We also focus on the model estimation from images of multi-dots. The impacts of the distance between dots on the estimator accuracy is studied as well as the precision gain of the ML estimator obtained by images of many separated dots.

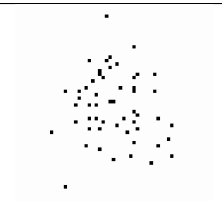
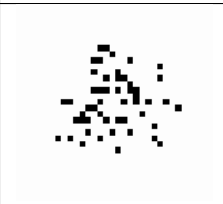
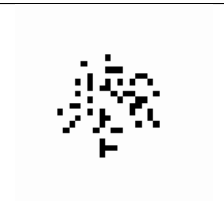
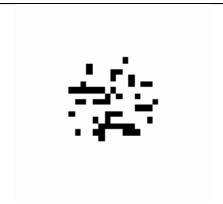


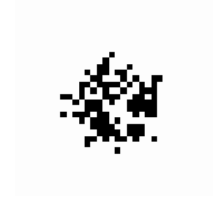


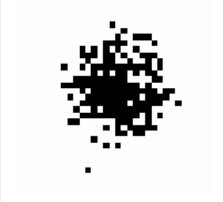
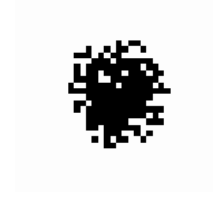
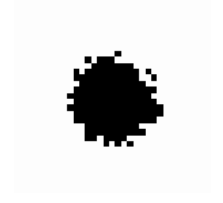
n	β			
	0.8	2	3	5
50				
100				
165				

Table 5.1.: The simulated dots with different values of parameters.

The impact of parameter β

To begin, various values of β with fixed other parameters are chosen to generate the data. The simulated images are generated with a single dot, each image has a size of 50×50 . The true centers μ of the dot is $(25, 25)$. With $\sigma^2 = 30, n = 95$, along with eight values of β , $\beta \in \{0.8, 1, 2, 3, 4, 5, 6, 7\}$; for each set of parameters, 1000 binary images are generated. For each image, Algorithm 5.1 is employed with $\epsilon = 0.1$. At total of $1000 \times 8 = 8000$ experiments were conducted.

The variances of the estimators and their Cramér-Rao lower bounds are shown in Fig. 5.5 and Fig. 5.6. The empirical variances of the estimators are asymptotic to the lower bounds, thanks to the efficiency of MLE [98, 42]. There are some values that are smaller (not much) than the lower bounds, there is no contradiction here because these are the **empirical variances** based on 1000 images, not the true variances of the ML estimators. Besides, the larger β , the further the variances from the lower bounds in case estimation of σ^2, β and n . The error in estimation of great β has been discussed with respect to KL distance, so that the accuracy of $\hat{\beta}$ also influences $\hat{\sigma}^2, \hat{n}$. Since the estimation of the location is mainly affected by the position of the black pixels, this effect on $\hat{\beta}$ is small on the estimation of the center μ . It is noted that the more scattered the particles are, the less accurate is the estimation of its center. Tab. 5.2 shows the bias of the estimators with respect to the different β .

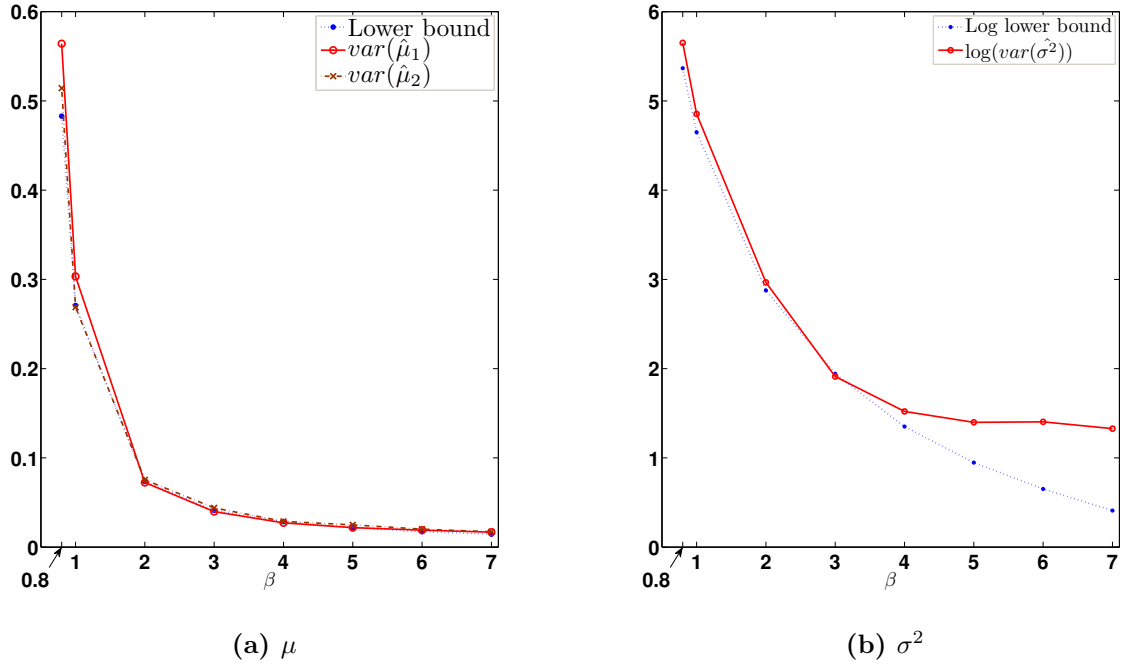


Figure 5.5.: The variance of the estimators vs. the lower bounds with respect to various β .

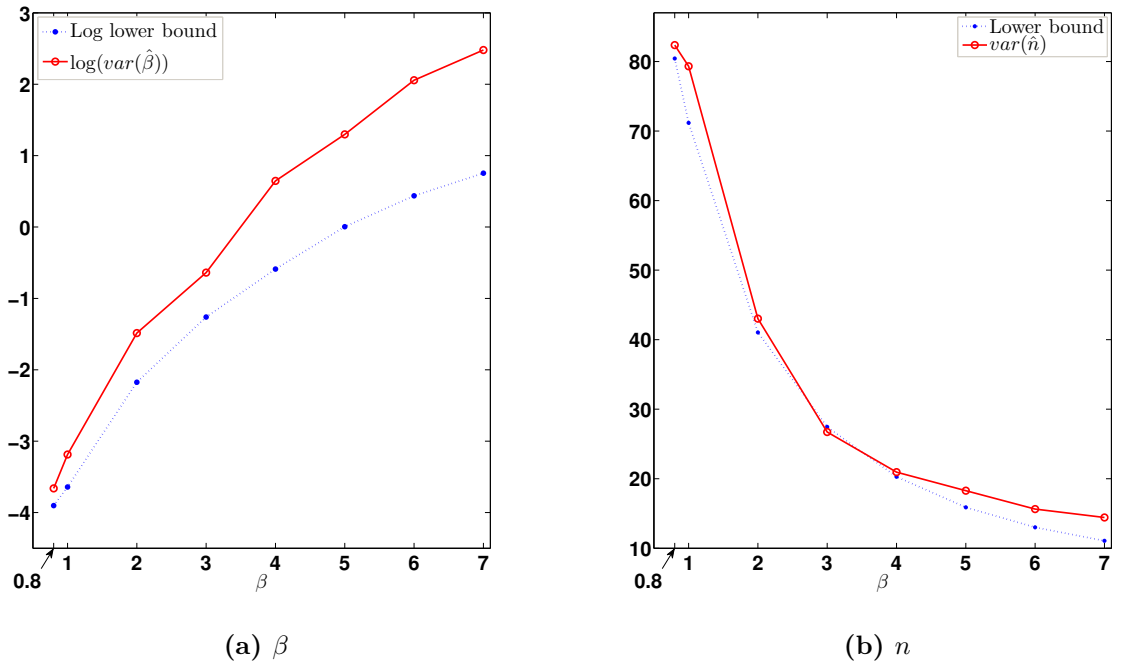


Figure 5.6.: The variance of the estimators vs. the lower bounds with respect to various β .

true β	0.8		2		3		5	
	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
σ^2	28.66	16.87	1.404	4.41	0.325	2.60	-0.509	2.01
β	0.266	0.16	0.207	0.48	0.277	0.73	0.100	1.91
n	0.7	9.08	0.251	6.56	-0.179	5.17	-0.114	4.28

Table 5.2.: Bias and standard deviation of the estimators, $\sigma^2 = 30, n = 95$.

The impact of parameter σ^2

The variation of σ^2 also affects the performance of the estimation. Tab. 5.3 illustrates the numerical results. One simulated dot data have been also conducted, the centers is at $(25, 25)$, $\beta = 2, n = 115$, along with three values of σ^2 , $\sigma^2 \in \{30, 50, 70\}$; for each set of parameters, 1000 binary images were generated. As shown in the table, the accuracy of $\hat{\sigma}^2$ decreases with large values of σ^2 , the same conclusion goes with $\hat{\beta}$. On the other hand, the estimator \hat{n} is less biased with large σ^2 but its variance gets bigger. Since σ^2 is large, the shape of the dots are less distinguishable, this makes the estimators less accurate with large σ^2 . When considering the relative errors, \hat{n} yields the smallest error, and $\hat{\beta}$ is less accurate.

true σ^2	30		50		70	
	Bias	Std.	Bias	Std.	Bias	Std.
σ^2	2.709	2.707	5.856	7.04	10.065	10.954
β	0.271	0.305	0.383	0.437	0.464	0.681
n	3.768	6.75	0.792	8.479	0.489	9.230

Table 5.3.: The bias and standard deviation with various σ^2 , $\beta = 2, n = 115$.

The impact of parameter n

The performance of the estimators with regard to n . The simulation of one dot were conducted, the centers is at $(25, 25)$, $\beta = 2, \sigma^2 = 50$, along with three values of n , $n \in \{50, 115, 165\}$; for each set of parameters, 1000 binary images were generated. As shown in Tab. 5.4, the accuracy of $\hat{\sigma}^2$ and $\hat{\beta}$ increase with large values of n . The relative errors of \hat{n} is also smaller with larger value of n . The accuracy is better with large n is already mentioned in the Fisher information matrix analysis (5.18), the black pixels contribute mainly to the amount of the information for the estimation.

true n	50		115		165	
	Bias	Std.	Bias	Std.	Bias	Std.
σ^2	10.00	11.15	5.856	7.04	3.94	5.28
β	1.09	2.07	0.383	0.437	0.23	0.30
n	0.28	6.61	0.792	8.479	1.77	8.25

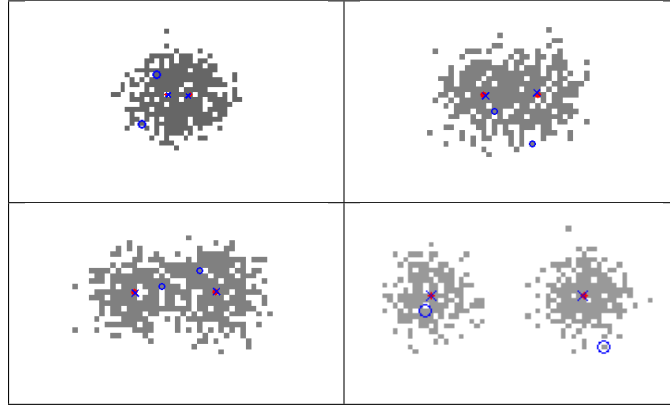
Table 5.4.: The bias and standard deviation with various value of n , $\beta = 2, \sigma^2 = 50$.

The impact of the distance between dots

The performance also depends on the distance between dots, Tab. 5.5 shows the results in four different distances of two centers, from very close to almost separated. As we can see, for the three last cases, two dots are not difficult to guess visually. But in the closest case, it is impossible to think that there are two dots in this image. The experiment is set up with 100 simulated images for each case, the parameters are $\sigma^2 = 50, \beta = 2, n = 165$. This result means that the model can discriminate efficiently two close cluster-dots as well as accurately determine the centers. We should emphasize that even in the closest distance case, when the discrimination between two dots is almost undetected, the algorithm can still have success in locating the centers. This feature of the model has a potential application in image recognition community. Moreover, when the dots are more separated, the estimation is more efficient with smaller variance. The distance between estimated centers and true ones is defined as

$$d(\hat{\mu}, \mu) = \sqrt{\sum_{k=1}^K \|\hat{\mu}_k - \mu_k\|_2^2}. \quad (5.36)$$

(a) The true centers (red dots), initial values (circle), and estimated centers (\mathbf{x})



(b)

Distance between 2 dots	5		10		15		30	
$d(\hat{\mu}, \mu)$	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
	0.52	0.26	0.53	0.48	0.41	0.14	0.34	0.137
σ^2	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
	5.74	6.27	3.21	5.81	2.94	3.85	2.48	3.92
β	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
	0.30	0.33	0.16	0.27	0.19	0.23	0.17	0.24
n	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
	0.41	6.37	2.24	7.10	-0.21	5.49	0.75	5.84

Table 5.5.: Two simulated dots with 4 different relative positions (a) and the error of the estimators for each case (b).

The performance with many dots

The impact of the number of dots, K , is also investigated. The experiment is conducted with $\sigma^2 = 50, \beta = 2, n = 165$, and all generated dots are completely separated. As shown in Tab. 5.6, with respect to K , more dots give better estimators. As mentioned in the analyst of σ^2 , with one dot, the errors of $\hat{\sigma}^2$ and \hat{n} are acceptable, while the relative error of $\hat{\beta}$ is worst. Since the accuracy of $\hat{\beta}$ is essential in authentication, improving its accuracy is important. Therefore, in order to have a good estimator of β , analyzing many dots is a good proposal. One reason of this behavior is that more dots will give more information about the shape, which are determined by the parameters of the model.

	1 dot		2 dots		4 dots		7 dots	
$d(\hat{\mu}, \mu)$	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
	0.33	0.18	0.37	0.12	0.35	0.083	0.36	0.07
σ^2	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
	3.25	5.35	2.71	3.64	1.53	2.53	1.33	2.16
β	0.19	0.32	0.17	0.21	0.09	0.13	0.085	0.11
n	1.18	8.51	0.92	6.14	0.56	3.99	0.79	3.01

Table 5.6.: The bias and standard deviation of the estimators with different number of cluster-dots K , $\sigma^2 = 50, \beta = 2, n = 165$ for each dot.

The performance with two randomly assigned dots

Next, the performances of the identification algorithm are assessed with the simulation of $K = 2$ dots. The true centers are not fixed but assigned randomly and uniformly, which means the positions of the centers can be either close or far from each other. The experiment intends to illustrate the performance of the estimation regardless of the location of the centers. This experiment also evaluates the accuracy of the estimation of the centers and the impact of n on the estimation. There are $N = 100$ simulations of 100×100 images of 2 dots obtained with $\sigma^2 = 50, \beta \in \{0.8, 2, 3, 5\}$, and $n \in \{50, 100, 165\}$. Tab. 5.1 visualizes the simulated dots with these values of the parameters. A total of $100 \times 3 \times 4 = 1200$ experiments have been conducted.

The analysis is executed through the bias and the standard deviation of the estimators. Tab. 5.7 and Tab. 5.8 show respectively the performances of the location estimators and $\hat{\sigma}^2$. As expected, estimators are more accurate with higher value of β and a large expected number of particles n . It also restates that the estimation with small values of β , close to zeros, is really erroneous. Moreover, it shows that the estimation of the centers is very good in the experiment. Firstly, because particles are distributed more uniformly with larger β (see Fig. 5.2). Secondly, when the mean number of particles is higher, the dots are more distinguishable from the background (white pixels).

About the performances of the estimator of β given in table Tab. 5.9, the best accuracy is obtained for the largest n , not only the bias but also the standard

n	50		100		165	
β	Avg.	Std.	Avg.	Std.	Avg.	Std.
0.8	2.51	1.79	1.55	0.65	1.14	0.48
2	0.82	0.33	0.55	0.26	0.39	0.24
3	0.63	0.40	0.40	0.15	0.31	0.26
5	0.51	0.33	0.34	0.38	0.21	0.11

Table 5.7.: The average and standard deviation of the distances between estimated centers and true ones.

n	50		100		165	
β	Bias	Std.	Bias	Std.	Bias	Std.
0.8	76.14	70.52	47.08	22.40	54.75	11.16
2	6.72	11.54	2.36	6.49	1.34	3.93
3	2.03	7.13	0.66	3.85	0.41	2.09
5	0.31	4.84	0.26	3.03	-0.05	1.98

Table 5.8.: The bias and standard deviation with various value of σ^2 .

deviation values. This impact of n has been clearly explained by the significant information of the black pixels. The behavior of the estimators according to β is also appropriate to the aforementioned analysis.

β	0.8		2		3		5	
n	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
50	0.44	0.36	0.74	1.98	0.90	2.28	2.64	5.55
100	0.26	0.11	0.16	0.34	0.21	0.55	0.71	2.50
165	0.32	0.05	0.08	0.202	0.05	0.34	-0.50	0.72

Table 5.9.: The bias and standard deviation with various value of β .

5.4.2. Bayesian estimation method

The performance of the Bayesian estimators is analyzed similarly to the maximum likelihood algorithm ones. Because they depends on the burn-in period, the algorithm setup is based on the analysis of the generated Markov Chains. We study the accuracy of the algorithm from simulated images of one dot. The experiments of two dots, or many dots are not conducted, since the behaviors of the estimation are similar to the aforementioned analysis with MLE.

The simulated images are generated with a single dot, the image have a size of 50×50 and the true centers μ is at (25,25). As said in sec.5.4.2, the priors are fixed, the expected values of the prior are set by the initial values, and the variance are chosen large in order to boost the impact of the likelihood in the acceptance ratio. The estimation is implemented with 10000 MCMC iterations and a burn-in period of 3000, which means 7000 last samples are used to estimate the parameters

over 10000 generated samples. Note that the burn-in period 3000 is based on the observation of the Markov chains of all experimented cases.

Performance of the estimator with regard to β

To begin, various values of β with fixed other parameters are chosen to generate the data. The parameter $\sigma^2 = 30$, along with $\beta \in \{0.8, 2, 3, 5\}$. As remarked, parameter η can be calculated from n when other parameters are known and vice versa. Then we generate the image with $n = 95$. Tab.5.10 shows the bias and the standard deviation of the generated samples for the estimation, the behavior of the standard deviation values is similar to the MLE, greater value of β is, smaller variances of $\hat{\sigma}^2$ and \hat{n} are. The variance of $\hat{\beta}$ is also greater with large β . Moreover, when comparing the standard deviation between two methods, the MCMC algorithm generally gives better results.

β	0.8		2		3		5	
	Bias	std.	Bias	std.	Bias	std.	Bias	std.
σ^2	2.38	10.81	-0.38	5.31	-2.28	2.13	-2.54	1.26
β	0.12	0.13	0.06	0.48	-0.50	0.33	-0.02	0.83
n	-2.13	8.70	-8.21	6.34	1.43	5.13	-9.43	3.99

Table 5.10.: The bias and standard deviation of the estimators with various β .

Performance of the estimator with regard to σ^2

The affects of variation of σ^2 on the performance of the estimation are discussed in this part. The fixed parameters are $\beta = 2$, $n = 115$ and various values of σ^2 , $\sigma^2 \in \{30, 50, 70\}$. The results in Tab.5.11 again confirm the less accuracy of the estimation when σ^2 is large. This behavior is similar to MLE.

σ^2	30		50		70	
	Bias	std.	Bias	std.	Bias	std.
σ^2	-0.45	2.68	-0.43	6.12	-3.42	10.22
β	-0.22	0.18	0.24	0.38	0.20	0.42
n	-1.51	6.53	4.27	7.49	-0.42	8.72

Table 5.11.: The bias and standard deviation of the estimators with various σ^2 .

Performance of the estimator with regard to n

The performance of the estimators with regard to n . The simulation of one dot were conducted with $\beta = 2, \sigma^2 = 50$, along with three values of n , $n \in \{50, 115, 165\}$. The estimators have smaller variance with large n , which means the black pixels give a significant information to the estimation procedure.

From these experiments, when we compare with MLE, the Metropolis-Hasting within Gibbs gives a slightly better performance, the variance values are generally

n	50		115		165	
	Bias	std.	Bias	std.	Bias	std.
σ^2	-12.68	9.74	-0.43	6.12	11.19	5.61
β	-0.32	0.41	0.24	0.38	0.65	0.40
n	2.59	6.28	4.27	7.49	8.53	8.36

Table 5.12.: The bias and standard deviation of the estimators with various n .

smaller when using MCMC. However, it cannot be said that MCMC is better than MLE. Since the fundamental difference between two approaches, MLE is a deterministic method while MCMC considers the parameters as random values, it means that MCMC generally generates more samples which makes smaller variance.

5.4.3. Estimation result from printed dots

The experiment is also executed on the real printed images. The dots from four popular technologies are taken, conventional offset, waterless offset, laser and inkjet printings. The resolution of each printing process for each sample are respectively 1200 dpi with offset printers, 600 dpi with laser printer, and 720 dpi with inkjet one. For each printer, 10 images of one dot are collected for the experiment. Tab.5.15 represents the images of the real dots and the realization generated from the estimated parameters. The maximum likelihood estimation and MCMC methods are applied in the estimation. Metropolis-Hasting within Gibbs algorithm is conducted with 20000 MCMC iterations with 5000 of burn-in period. Tab. 5.13 and Tab. 5.14 give the numerical values of the estimators by two numerical estimation methods as well as the standard deviation values.

	Conventional Offset		Waterless offset		Laser		Inkjet	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
σ^2	25.10	8.74	165.28	60.52	128.91	48.56	723.25	132.57
β	2.93	0.46	2.81	0.76	1.58	0.58	3.12	0.39
n	89.10	29.87	533.38	148.322	334.42	117.64	2546.1	471.39

Table 5.13.: The estimators of the real dots by MLE.

	Conventional Offset		Waterless offset		Laser		Inkjet	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
σ^2	24.88	9.18	126.42	40.41	84.32	73.77	705.40	138.72
β	3.47	1.28	4.85	3.84	1.17	0.56	3.36	0.77
n	84.10	29.33	469.55	115.07	324.46	105.29	2468.5	459.98

Table 5.14.: The estimators of the real dots by MCMC.

We can see that the two methods do not provide important differences with respect to estimators. When we look at the realizations, the visual results are really close to each others. With the real data, the generated dots technology are mostly

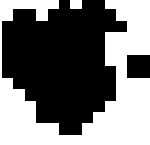
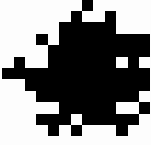
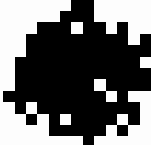


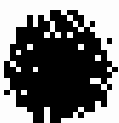






	Real	MLE	MCMC
Conventional Offset			
Waterless Offset			
Laser			
Inkjet			

Table 5.15.: The estimation from four printing processes with two estimation methods.

fragmentary, specially on the borders. Considering the dimensionless parameter β , the variances from ML estimators are likely smaller than the Bayesian ones. As analyzed, with great values of β , the realizations are less likely to be distinguishable, which means the MCMC samples of β are going to jump more randomly and further than in the case of small β .

5.5. Conclusion

In the chapter, the spatial binary model for the image of the cluster-dots is introduced and rigorously analyzed. The role of each parameter that contribute to the formation of the cluster-dots as well as the properties of the model are deeply presented. The model has been shown to have a potential ideal to realize the particular shape of the printed dots, and in fact, the model seems adequate with real dots, see Tab. 5.1. As the model is proposed, the estimation methods are developed. Our approaches come both from frequentist point of view and a Bayesian one. The ef-

fects of the values of the parameters on the accuracy of the estimators are explained. Through the experiment, both estimation approaches perform well, specially with the simulated data. With MLE, there is a drawback since it is very time consuming when there are many dots, since the size of the Hessian matrix is increased. With MCMC method, we only have to increase the number of priors, which is less complex. Therefore, it would be wiser to use MCMC when dealing with many dots. As shown in sec. 5.4.3, the estimation should be more similar to the real data. This is due to the model which seems to be too “rough”.

In fact, the model considers the pixels as independent, which is not enough for real-world images. In the next chapter, the interaction of the neighborhood will be studied in order to fulfill the model of the formation of the cluster-dots.

Chapter 6.

Markov spatial binary model

In this chapter, we introduce another binary response model which takes into account the effect of the vicinity of a current pixel. As demonstrated in the previous chapter, the interaction of the geometric vicinity is a natural phenomenon which cannot be ignored, the question is how close a model can adapt that reality. In chapter 5, only the distances between the pixel and the centers are taken into account. Instead of considering that the pixels are independent as in chapter 5, the model is developed with a new interactive parameter that quantifies the influence of the vicinity of the considered pixel. After the construction of the model and its properties, the estimation method is also established with a Bayesian estimation method.

The chapter is constructed with three main sections. First section focuses on constructing the model according to the observations. Second section describes the estimation method for the parameters and the algorithm. The third one is used to validate the model by analyzing the simulated images as well as real printed dots, the accuracy of the estimation algorithm is also taken into account. A concluding section then ends the chapter.

6.1. Interaction model with Markov chain

Recall the desirable properties of the model mentioned in chapter 3, the emergence of the black pixels are lower when they are further from the centers, and the interaction between pixels. In fact, when a pixel is black, its neighbor(s) is(are) more likely blackened. Moreover, when the pixel is further from the centers of the dots, this influence of the black pixel(s) in its vicinity is also smaller. Taking into account this phenomena would make the dot less fragmented, which makes the model closer to the realistic image and more precise. Therefore, the probability of a current pixel is also influenced by the realization of its neighbor(s). Let V_s be a neighborhood of the pixel s , the geometric shape of V_s is independent of $s \in S$. The interaction of the neighbor(s) can be modeled by the Markov random field \mathbf{U} :

$$P(U_s = u_s | (U_t = u_t)_{t \neq s}) = P(U_s = u_s | (U_t = u_t)_{t \in V_s}), \quad (6.1)$$

i.e. the probability that the pixel s takes the value u_s conditional on all other pixels of the image equals to the probability of u_s conditional on the values of the pixels in the neighborhood V_s . This probability also has to depend on the distance of the pixel s to the center, which means $P(U_s = 0 | (U_t = u_t)_{t \neq s})$ decreases when the pixels s goes further from the center, i.e. $\|s - \mu\|$ increases.

For simplicity, we will restrict the neighborhood to one pixel only. In the other words, the interactions are taken into account by modeling \mathbf{U} by a Markov chain. Considering the image with N pixels, the realization of the binary image is described as follow

$$P(\mathbf{U} = \mathbf{u}) = P(U_1 = u_1) \prod_{i=2}^N P(U_i = u_i \mid U_{i-1} = u_{i-1}). \quad (6.2)$$

A 2D image is always able to be transformed into a one-dimensional chain by scanning the pixels line by line or column by column (see Fig. 6.1). However, thanks to the geometric property of the Hilbert curve [99] (see Fig. 6.1c), it is suitable to index the sequence $\{U_s\}$ along this Hilbert path. The Markov chain with Hilbert path is used in many image processing applications [100, 101]. Moreover, Pseudo-Hilbert Scan algorithm can be applied to arbitrarily-sized image [102].

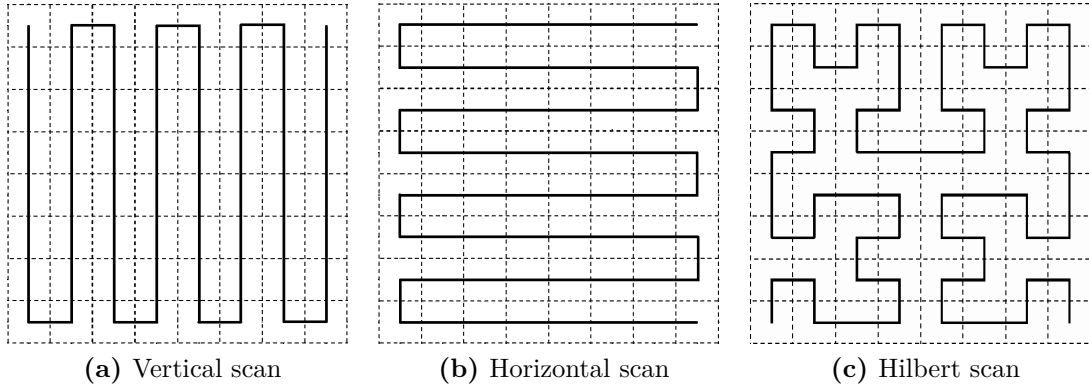


Figure 6.1.: Scanning of 8×8 image.

The Markov chain (6.2) encompasses the parameters of the first probability

$$P(U_1 = 0), \quad (6.3)$$

and transition probabilities

$$P(U_i = 0 \mid U_{i-1} = u), \quad (6.4)$$

for $1 < i \leq N$. From the original idea in chapter 5, we introduce parameter λ and propose the parametric probabilities for (6.3), (6.4). The ability to blacken the surrounding pixels of the k -th dot is still measured by a function $p_k(\cdot)$. As mentioned, these probabilities are inversely proportional to the distance $\|s - \mu_k\|$ and its support is the subset of $(0, 1]$,

$$P(U_1 = 0) = 1 - \prod_{k=1}^K (1 - p_k(s_1)) \quad (6.5)$$

$$P(U_i = 0 \mid U_{i-1} = u) = 1 - \prod_{k=1}^K (1 - p_k(s_i)) \lambda^{(1-u)}. \quad (6.6)$$

Consider the neighbor is black, $U_{i-1} = 0$, the new positive parameter λ plays a part in the influence of the neighbor of the current pixel. With $\lambda = 1$, the probabilities (6.6) only depend on the relative position to the centers, that is the model described in chapter 5, $P(U_i | U_{i-1}) = P(U_i)$. When $\lambda < 1$, the smaller λ is the more likely black the current pixel is. In fact, it can be seen from (6.6) that $P(U_i = 0 | U_{i-1} = 0)$ is greater with smaller λ . When the neighbor pixel is white, $U_{i-1} = 1$, the probabilities depend on the distances to the centers of the current pixel. With these properties, the dots will be more compact, and the density of black pixels at the areas closer to the centers will be higher.

The transition probabilities depend on $p_k(s)$ which relates to the position of the pixels, this implies that \mathbf{U} is a non-stationary Markov chain. The expected number of black pixels of the image n_b is calculated as

$$n_b = \sum_{i=1}^N E(\delta[U_i]) = \sum_{i=1}^N P(U_i = 0), \quad (6.7)$$

where $P(U_1 = 0)$ is computed by (6.5), while the other terms is calculated by (6.6) with the recurrence relation and the law of total probability

$$P(U_i = 0) = P(U_i = 0 | U_{i-1} = 0)P(U_{i-1} = 0) + P(U_i = 0 | U_{i-1} = 1)P(U_{i-1} = 1), \quad (6.8)$$

with $1 < i \leq N$. Let us recall the relation $P(U_i = 0) + P(U_i = 1) = 1$ for all i . It is noted that with smaller λ , $P(U_i = 0 | U_{i-1} = 0)$ is greater, that means $P(U_i = 0)$ also becomes greater. This implies that smaller λ will increase the number of black pixels. In the following, the Gaussian power kernel (5.11) is chosen for the form of p_k ,

$$p_k(s) = \eta \exp \left\{ -\frac{1}{2} \left(\frac{\|s - \mu_k\|_2^2}{\sigma^2} \right)^\beta \right\}. \quad (6.9)$$

Fig. 6.2 illustrates the simulation of one dot with σ^2 , β equal to 70, 5 respectively and various values of λ, η . As we can see, the dots with small λ are more “connected”, η controls the amplitude of the kernel, the density of black pixels increases with η . When η is large, the density of black pixels near the center with large λ is not different when λ is small.

6.2. Estimation

The frequentist estimation such as MLE is not mentioned because of the high dimension of the vector of parameters which increases significantly the complexity of the computation.

In this model, the method estimation of the parameters is based on Bayesian inference. Let us recall our problem, the interested parameters in (6.5), (6.6) are $\theta_{\mathbf{U}} = [\{\mu_k\}, \sigma^2, \beta, \eta, \lambda]$. The method of Minimum Mean Squared Error is described,

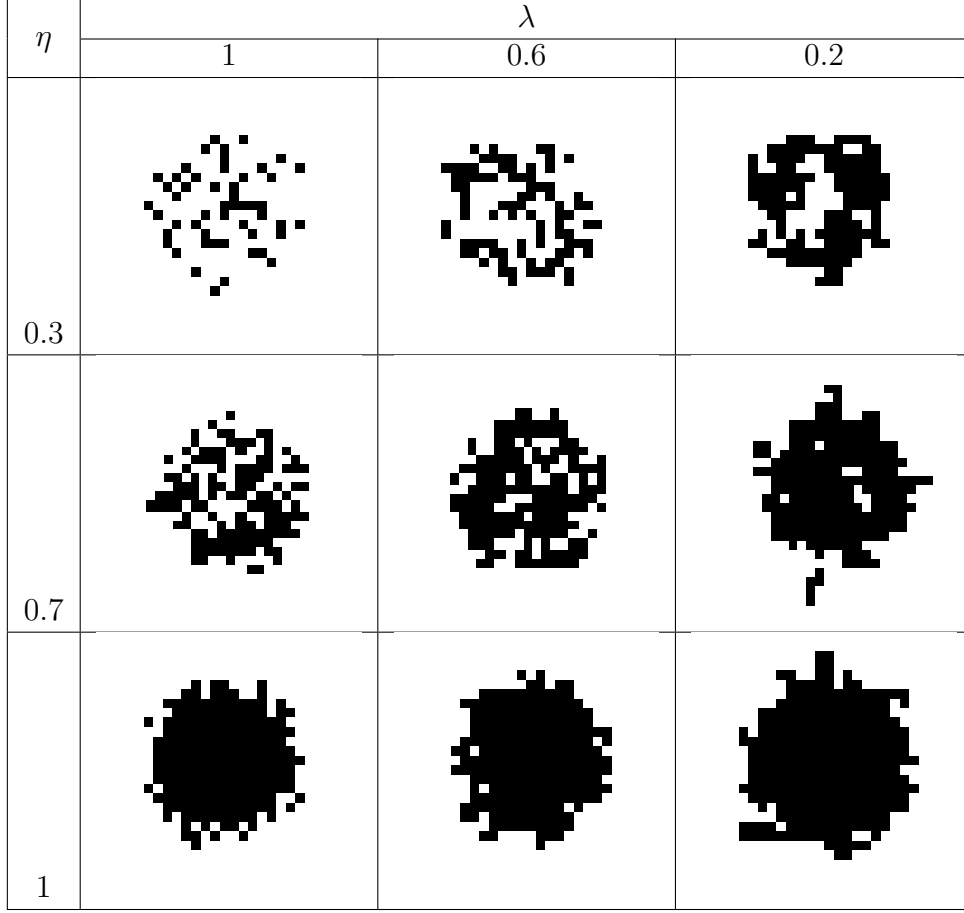


Figure 6.2.: The simulation of the model (6.5), (6.6) with Gaussian power kernel with $\sigma^2 = 70$, $\beta = 5$ and different values of λ, η .

the estimator $\hat{\theta}$ of $\theta_{\mathbf{U}}$ given the observed data \mathbf{u} is

$$\hat{\theta}(\mathbf{u}) = \int \theta p(\theta | \mathbf{u}) d\theta \quad (6.10)$$

where $p(\theta | \mathbf{u})$ is the posterior distribution. To approximate the integral (6.10), MCMC algorithm is applied. Since the vector of parameters is high-dimension, it is more convenient to use Gibbs sampling algorithm. In particular, the estimation is based on Metropolis-Hastings within Gibbs algorithm Algorithm 2.5. Moreover, the interested parameters are considered to be independent. The chosen strategy is to run a random walk (2.24) to generate the Markov chain. Particularly, if the normal distribution is used, then the proposal function is symmetric. Similar to sec. 5.4.2, we generate $\theta_2 = \log(\sigma^2), \theta_3 = \log(\beta)$ instead of σ^2, β ; the transformation (5.34) is applied for η and λ to generate θ_4, θ_5 respectively, i.e. $\eta = \frac{1}{1+e^{\theta_4}}, \lambda = \frac{1}{1+e^{\theta_5}}$. For convenient notation, we rewrite the vector of interested parameters as $\theta = [\theta_1, \theta_2, \theta_3, \theta_4, \theta_5]$. The Metropolis-Hastings within Gibbs using random walk is summarized in Algorithm 6.1.

Recall that $\theta_{-b}^{(i-1)} = [\theta_1^{(i)}, \dots, \theta_{b-1}^{(i)}, \theta_{b+1}^{(i-1)}, \dots, \theta_5^{(i-1)}]$. The initialization θ is obtained by the careful seeding, as in sec. 5.1, for the means. The other parameters

Algorithm 6.1 Sampling algorithm with random walk.

Initialize $\theta^{(0)} = [\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)}, \theta_4^{(0)}, \theta_5^{(0)}]$
for $i = 1, \dots, N_iter$ **do**

 for $b = 1, \dots, 5$ **do**

 Generate $\theta_b^* \sim \mathcal{N}(\theta_b^{(i-1)}, \sigma_b^2)$

Compute the acceptance ratio

$$\alpha(\theta_b^{(i-1)}, \theta_b^*) = \min \left\{ \frac{p(\theta_b^* | \theta_{-b}^{(i-1)}, \mathbf{u})}{p(\theta_b^{(i-1)} | \theta_{-b}^{(i-1)}, \mathbf{u})}, 1 \right\}$$

Decide

$$\theta_b^{(i)} = \begin{cases} \theta_b^* & \text{with probability } \alpha(\theta_b^{(i-1)}, \theta_b^*) \\ \theta_b^{(i-1)} & \text{otherwise} \end{cases}$$

end for
end for

are respectively initiated as $\sigma^2 = 0.6 \times \frac{N}{\pi K}$, $\beta = 2$. Since η and λ are in $(0, 1]$, intuitively, we may set both of them equal to 0.51. The prior distributions are chosen with fixed parameters by experience. Each center $\{\mu_k\}$ has **multivariate normal** distribution, σ^2 distributes with **log-normal law**, β has **gamma law**, while both η and λ are **beta-distributed**. From Bayes's law, we have

$$p(\theta_b | \theta_{-b}, \mathbf{u}) = \frac{p(\mathbf{u} | \theta_b, \theta_{-b}) p(\theta_b, \theta_{-b})}{p(\theta_{-b}, \mathbf{u})}$$

where $p(\mathbf{u} | \theta_b, \theta_{-b})$ is the likelihood and $p(\theta_b, \theta_{-b})$ is the prior. Since each parameter is distributed independently, the acceptance ratio in Algorithm 6.1 is rewritten as

$$\alpha(\theta_b^{(i-1)}, \theta_b^*) = \min \left\{ \frac{p(\mathbf{u} | \theta_b^*, \theta_{-b}) p_b(\theta_b^*)}{p(\mathbf{u} | \theta_b^{(i-1)}, \theta_{-b}) p_b(\theta_b^{(i-1)})}, 1 \right\}. \quad (6.11)$$

with p_b the prior of parameter θ_b . Let us note that when the priors are uniform or when the variance of the priors are large, the acceptance ratio mainly depends on the likelihood ratio. The choice of the variances of the priors depends on the confidence of the initialization and the experience.

6.3. Experimental results and performance

In this section, we aim at assessing both the performance of the estimators and the adequacy of the model. The Metropolis-Hasting within Gibbs algorithm is analyzed, the performance are evaluated from the images generated by the parametric model.

The simulated images are generated with a size of 64×64 . As mentioned, the priors are fixed, the expected values of the prior are set by the initial values, and

the variance are chosen large in order to boost the impact of the likelihood in the acceptance ratio. The estimation is implemented with 20000 MCMC iterations and a burn-in period of 5000, which means 15000 last samples are used to estimate the parameters over 20000 generated samples. Note that the burn-in period 5000 is based on the observation of the Markov chains of all experimented cases.

Then we study the adequacy of the estimated model with few printings coming from various printers and technologies. The parametric model is justified by the images obtained from four printing technologies by the estimation method.

6.3.1. Numerical performance with simulated data

Performance of the estimator with respect to β

To begin, various values of β with fixed other parameters are chosen to generate the data. The simulated images are generated with a single dot. The true centers μ of the dot is $(32, 32)$. With $\sigma^2 = 50, \eta = 0.8, \lambda = 0.6$, along with three values of β , $\beta \in \{0.8, 2, 5\}$. Tab. 6.1 shows the bias and the standard deviation of the generated samples for the estimation. Observing the behavior of the standard deviation values, greater value of β is smaller variances of $\hat{\sigma}^2$ and $\hat{\eta}$ are. In contrast, the variance of $\hat{\beta}$ and $\hat{\lambda}$ are greater with large β . The reasons for this performance of $\hat{\sigma}^2, \hat{\eta}$ and $\hat{\beta}$ are explained in the previous chapter. About $\hat{\lambda}$, with a large value of β , the fragment mainly happens far from the center. It means that the information of λ is smaller, this makes λ more difficult to estimate. While with smaller β , the contribution of λ to the connected black pixels is more visible, then the samples are less spread out.

β	0.8		2		5	
	Bias	std.	Bias	std.	Bias	std.
σ^2	10.72	10.81	-2.96	6.03	-2.05	3.00
β	0.13	0.11	-0.41	0.25	-1.12	0.87
η	-0.12	0.09	-0.07	0.08	-0.01	0.06
λ	0.00	0.05	-0.02	0.08	-0.05	0.10

Table 6.1.: The bias, standard deviation with various $\beta, \sigma^2 = 50, \eta = 0.8, \lambda = 0.6$.

Performance of the estimator with respect to σ^2

The variation of σ^2 also affects the performance of the estimation. Tab. 6.2 illustrates the numerical results. One simulated dot data were also conducted, the centers are at $(32, 32)$, $\beta = 2, \eta = 0.8, \lambda = 0.6$, along with three values of σ^2 , $\sigma^2 \in \{30, 50, 70\}$. The results in Tab. 6.2 again confirm the smaller accuracy of $\hat{\sigma}^2$ when σ^2 is large. The standard deviation increases according to the values of σ^2 .

Performance of the estimator with respect to η

The performance of the estimators with regard to η is given in Tab. 6.3. The simulation of one dot were conducted with $\beta = 2, \sigma^2 = 50, \lambda = 0.6$, along with three

6.3 Experimental results and performance

σ^2	30		50		70	
	Bias	std.	Bias	std.	Bias	std.
σ^2	-2.89	3.61	-2.96	6.03	1.96	6.17
β	0.39	0.73	-0.41	0.25	-0.10	0.27
η	-0.05	0.10	-0.07	0.08	-0.02	0.06
λ	-0.03	0.10	-0.02	0.08	0.04	0.07

Table 6.2.: The bias, standard deviation with various σ^2 , $\beta = 2, \eta = 0.8, \lambda = 0.6$.

values of η , $\eta \in \{0.5, 0.8, 0.97\}$. The estimators of σ^2 have smaller variance with large η , which means the black pixels give a significant information to the estimation procedure. With $\hat{\eta}$ and $\hat{\lambda}$, the biases increase, this is due to the fact that when η is large, the role of λ is less important, it also happen with η . This phenomena can be seen in Fig. 6.2.

η	0.5		0.8		0.97	
	Bias	std.	Bias	std.	Bias	std.
σ^2	6.14	7.95	-2.96	6.03	4.18	5.69
β	-0.04	0.45	-0.41	0.25	0.09	0.42
η	0.01	0.08	-0.07	0.08	-0.16	0.08
λ	0.03	0.07	-0.02	0.08	-0.06	0.08

Table 6.3.: The bias and standard deviation with various η , $\beta = 2, \sigma^2 = 50, \lambda = 0.6$.

Performance of the estimator with respect to λ

The simulation of one dot was conducted with $\beta = 2, \sigma^2 = 50, \eta = 0.8$, along with three values of λ , $\lambda \in \{0.2, 0.6, 0.9\}$. Tab.6.4 shows the bias and the standard deviation of the generated samples for the estimation. The estimators of σ^2, β, η have decreasing variance when λ increases. We know that the number of black pixels decreases when the value of λ increase, and the dots are more fragmented. This makes the influences of σ^2, β, η clearer. On the other hand, the estimation of λ is better for small values, this is because the connection between black pixels has strongly emerged.

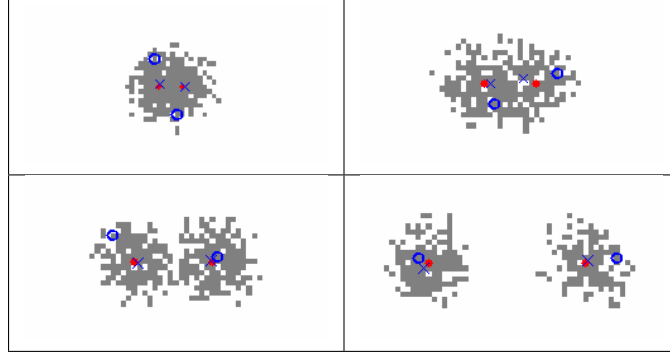
λ	0.2		0.6		0.9	
	Bias	std.	Bias	std.	Bias	std.
σ^2	0.54	8.11	-2.96	6.03	-2.33	5.58
β	-0.38	0.32	-0.41	0.25	-0.37	0.25
η	-0.12	0.11	-0.07	0.08	-0.08	0.07
λ	-0.01	0.04	-0.02	0.08	-0.10	0.09

Table 6.4.: The bias and standard deviation with various λ , $\beta = 2, \sigma^2 = 50, \eta = 0.8$.

The impact of the distance between dots

The results in four different distances of two centers are illustrated in Tab. 6.5, from very close to totally separated. As observed, two dots are not difficult to guess, visually, in three last cases. But in the closest case, it is impossible to visually distinguish two dots in this image. The experiment is conducted with the parameters $\sigma^2 = 50, \beta = 2, \eta = 0.8, \lambda = 0.6$. This result means that the interactive model can discriminate efficiently two close cluster-dots as well as accurately determine the centers. We should emphasize that even in the closest distance case, when the discrimination between two dots is almost undetected, the algorithm can still localize successfully the centers. This is an important feature of the model which has a potential application in image recognition community. In addition, the estimation is more efficient when the dots are further from each other. The distance between estimated centers and true ones is defined as (5.36).

(a) The true centers (red dots), initial values (circle), and estimated centers (\mathbf{x})



(b)

	5		10		15		30	
$d(\hat{\mu}, \mu)$	0.58		2.09		0.93		1.05	
	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
σ^2	-5.58	6.602	3.05	5.044	2.57	3.326	-0.44	3.081
β	-0.55	0.247	-0.09	0.302	0.25	0.271	-0.12	0.190
η	0.01	0.087	-0.07	0.069	-0.02	0.053	0.08	0.044
λ	-0.04	0.069	-0.11	0.066	-0.02	0.066	0.13	0.057

Table 6.5.: Two simulated dots with 4 different relative positions (a) and the estimated parameters for each case (b), $\sigma^2 = 50, \beta = 2, \eta = 0.8, \lambda = 0.6$.

The performance with many dots

The impact of the number of dots, K , is analyzed. The experiment is set up with $\sigma^2 = 50, \beta = 2, \eta = 0.8, \lambda = 0.6$, and all generated dots are completely separated. As seen in Tab. 6.6, with respect to K , the better estimators come with more dots. The accuracy of $\hat{\beta}$ is vital in authentication, it is important to improve its accuracy. Hence, analyzing many dots is a good mean to have a better estimator of β . The

reason of this behavior is that more dots will give more information about the shape, which are determined by the parameters of the model.

	1 dot		2 dots		4 dots		7 dots	
$d(\hat{\mu}, \mu)$	1.00		1.05		0.55		0.51	
	Bias	Std.	Bias	Std.	Bias	Std.	Bias	Std.
σ^2	-2.96	6.03	-0.44	3.08	1.37	2.42	-1.24	1.92
β	-0.41	0.25	-0.12	0.19	0.13	0.18	0.006	0.15
η	-0.07	0.08	0.08	0.04	0.004	0.04	0.03	0.03
λ	-0.02	0.08	0.13	0.06	0.045	0.05	0.013	0.03

Table 6.6.: The bias and standard deviation of the estimators with different number of cluster-dots K , $\sigma^2 = 50$, $\beta = 2$, $\eta = 0.8$, $\lambda = 0.6$.

6.3.2. Estimation result from printed dots

The experiment is also executed on the real printed images. The dots from four popular technologies are taken, conventional offset, waterless offset, laser and inkjet printings. The resolution of each printing process for each sample are respectively 1200 dpi with offset printers, 600 dpi with laser printer, and 720 dpi with inkjet one. For each printer, 10 images of one dot are collected for the experiment. Tab. 6.8 represents the images of the real dots and the realization generated from the estimated parameters. The MCMC method is applied in the estimation. Metropolis-Hasting within Gibbs algorithm is conducted with 30000 MCMC iterations with 10000 of burn-in period.

	Conventional Offset		Waterless offset		Laser		Inkjet	
	Avg.	Std.	Avg.	Std.	Avg.	Std.	Avg.	Std.
σ^2	28.80	9.2	139.83	42.71	98.54	49.37	261.73	53.54
β	3.71	1.4	5.86	3.61	1.21	0.92	2.28	0.51
η	0.96	0.05	0.99	0.0076	0.59	0.32	0.99	0.01
λ	0.67	0.15	0.46	0.29	0.30	0.06	0.20	0.04

Table 6.7.: The estimators of the real dots by MCMC.

Tab. 6.7 gives the numerical values of the estimators by two numerical estimation methods as well as the standard deviation values. Tab. 6.8 visually compares the printed dots and the realizations from the Markov spatial model and Spatial binary model in the previous chapter. It gives a very promising result when the interaction is taken into account. The first difference between two model is that the realization from Markov spatial model is more “connected” than the one from spatial binary model, so that the realization from Markov spatial model is closer to the real dot. With the offset and inkjet images, the fragment in the dots and near the border is handled better. The shape of the dot from inkjet printer is more complicated to simulate since the concave parts of it. About the print from laser printer, we can see the substantial improvement using the Markov spatial model w.r.t. the

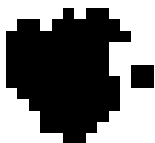
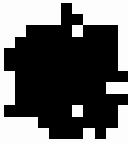
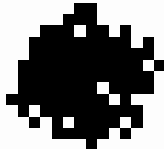

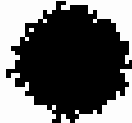
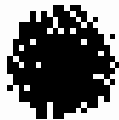



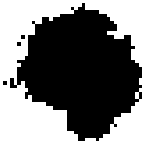
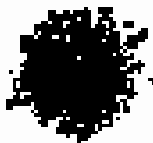
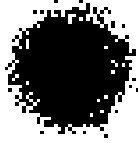
	Real	Markov spatial model	Spatial binary model
Conventional Offset			
Waterless Offset			
Laser			
Inkjet			

Table 6.8.: The estimation from three printing processes and the realizations from the models.

spatial binary model in chapter 5. The simulated dot is now much more realistic and compact. In sec. B.4, we give another estimation from a laser printed dots.

6.4. Conclusion

In this chapter, we have proposed an improved idea filling the gap in the spatial binary model in chapter 5 by considering the interactive phenomena. The new “interactive” parameter λ in the spatial binary model fulfills the required properties. This model successfully describes the formation of the shape of the binary dot. The spreading of the ink on the surface was described as the emergence of the black pixels and the interaction between pixels, they depend on the distance between the centers of the dots, and these features are modeled by non-stationary Markov random field.

An estimation method for the parameters is established, based on the Metropolis-Hasting within Gibbs algorithm. The performance of this numerical estimation is well validated, not only with the simulated data but also with the real dots from

four different printing technologies. Thanks to the non-stationary Markov random field, the realizations are very close to the real images. The fragmentation on the border of the dot is better treated, and the density of the black pixels is higher at the inked areas since the dots are more compact. In the other words, the simulated micro-print is now more “natural”.

Chapter 7.

Conclusions

Inspired by authentication using 2D graphical codes, increasing the resolution of the codes would increase the storage for the information and the security level. In order to develop the authentication for this high security code, interpreting the random behaviors as well as modeling these characteristics is essential. In this dissertation, we have analyzed and modeled the printing process from a physical and signal processing perspective. The stochastic behavior of the printing process at the microscopic scale has been investigated through digital gray images. The core of the model includes two parts, the first one is the gray scale model of the captured-printed code. The second part is the spatial binary model which simulates the shape of the microscopic prints.

Chapter 2 has been devoted to present briefly the commonly used printing technologies and the main elements of each has been presented. This helped us to have a general view of a printing process. We have reviewed some proposed printing models to see their applications and the limits of these models to our problem. We have also represented the important materials for parameters estimation that are used in the thesis.

It is necessary to interpret the behavior of ink spreading on the substrates in order to propose an accurate model. In chapter 3, we did an analysis on the real microscopic printed form. The observation focuses on analyzing the profile of a single dot and the degradation for different configurations. The experiment has pointed out that the distance from the center affects the ink spreading phenomenon, and it depends on the technology. For example the dot from a laser printer is made of many small particles, while the dot from offset printer is solid and more “connected”. Another characteristic for a print is the interaction between dots which cause the ink spreading on the overlap areas. This experiment gave us a physical overview on the spatial spreading of the ink on the substrates. Moreover, we have also analyzed the parameters of the geometrical features such as area, perimeter for each configuration and each printing process. This analysis helped us understand the difference of the print with different technologies, substrates.

Thanks to the properties obtained from the analysis of dots conducted in chapter 3, the idea for modeling the gray level and the shape of printed dot at the microscopic scale has been proposed in chapter 4. We formulated two parts of the model, the gray level of blank and printed areas, and the spatial spreading of binary process. This chapter focused on finding the parametric model and the estimation of the gray level

Conclusions

that takes into account the specific shapes of the blank and printed distributions. There are few threshold models developed to create the binary images from gray scale images [80, 81, 82], but none of these studies takes into account the printing technology and the substrate. The most common used distribution is Gaussian distribution which cannot handle the asymmetric histograms. The trials on other parametric distribution, therefore, have been conducted. We used Kolmogorov-Smirnov distance as a criterion to measure the goodness-of-fit of the parametric distributions. The test was conducted on the samples from four commonly used printing technologies and two kinds of paper along with five parametric distributions. These five parametric distributions were chosen by the analysis of the moments of the data with the help of the Pearson diagram in Fig.4.2. In order to have consistent results, we have shown that the physical conditions in the acquisition must be carefully preserved.

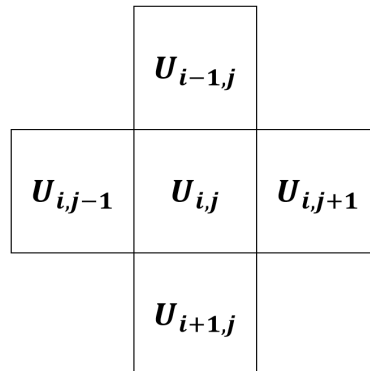
For the blank area, the best candidate is **Beta** distribution with the smallest K-S distance. Depending on the specific technology, printer and the substrate, the experiment results two good options which are **Inverse-gamma** and **Log-normal** distributions for the gray scale model. Indeed, the results from these two distributions are quite close and a lot better than the others. When the sum of K-S distances are compared, the **Inverse-gamma** is the most appropriate distribution. Another remark is that even with the same technology but different printers, the results can be different (see Tab.4.6). Final conclusion from this experiment is that the samples from the same printer and material for a short period of time, for example three weeks in our case, are consistent. The substrates also make a difference, for instance with the inkjet printer, the log-normal distribution is the better proposal for the print on uncoated paper; for coated paper, inverse-gamma distribution gives a better fit.

Chapter 5 and chapter 6 concentrated on modeling the spatial binary image. From the analysis of microscopic prints conducted in chapter 3, in chapter 5, we proposed a probabilistic model which considers the pixels of the image as a sequence of independent random variables having Bernoulli distribution with the parameter is the probability of being blackened of a pixel. This probability varies according to the distance to the centers of the dots. Each dot has different impact to this probability according to the distance to the center of the dot to that pixel. In the model, the parameters that contribute to the formation of the cluster-dots as well as the properties of the model are introduced and rigorously analyzed. Since the shape of the cluster-dots is very random, the choice of parameters that can encompass this wide variety of form is really challenging. From the analysis of dots conducted in chapter 3 and a few attempts, the Gaussian power kernel was chosen for $p_k(s)$ with the shape parameter β . The kernel encompasses many shape, from the Laplace shape to the uniform shape. The estimation for the parameters was developed based on Maximum likelihood estimation and MCMC methods. Since the mathematical form of the model is not simple, the solution of Maximum likelihood estimation method has to be obtained by the numerical estimation, particularly the augmented Lagrangian method. The MCMC method used is Metropolis-Hasting within Gibbs algorithm. The model was validated by the simulated data as well as the real dots

from four printing technologies. We have analyzed the influence of the variation of each parameter to the accuracy of the algorithms. The experiment we conducted with two dots indicates that when the dots are more separated the estimation is more efficient. Moreover, even with a very close distance, the model can still discriminate efficiently two cluster-dots as well as accurately determines the centers (see Tab. 5.5). Estimating many dots also gives better estimators than fewer dots. Although the model seems adequate with real dots, the realizations of the model are still less “connected” than the real images. This is because the model considers the pixels in the image as independent.

Therefore, an improved idea was proposed and developed in chapter 6. The model in chapter 5 satisfies the property that the pixels further from the centers will have lower probability of being black. With the Markov spatial binary model, not only this property is fulfilled but also the spatial interaction between pixels is taken into account. However, the classical stationary Markov chain is not appropriate. Based on the analysis of the dots in chapter 3, this spatial interactive feature also varies according the distance from the centers of the dots. Since the Markov process has to be non-stationary, this makes the model mathematically more complicated. These are two fundamental properties for the formation of the shape of the cluster-dots, and they are well modeled in our Markov spatial binary model. The estimation for the parameters of the Markov spatial binary model is well established based on MCMC method. Particularly, Metropolis-Hasting within Gibbs algorithm is developed for the estimation. The model was also validated with both simulated data and real dots from four different printing processes. The experiment with the simulated data has been conducted similarly to what we have done in chapter 5, and the behaviors of the estimators are the same. The realizations of the model are very close to the real micro-dots images. The success in simulating the microscopic prints enables us to improve the quality in quantifying the distortion between the original document and a counterfeit, increasing the security level of the graphical codes.

Finally, we would like to indicate some possible approaches for the future researches. In chapter 6, we have analyzed the spatial interaction with “one-neighbor” vicinity, i.e. U_{i-1} is the only neighbor of U_i . We can expand the model with a more efficient vicinity, for example four-nearest-neighbor is a well-known option. A sort of spatial situations that are worth to discuss had been described in the statistical analysis of lattice systems of Besag [103].



Conclusions

Other extension for our models occurs in the form of the kernel. Throughout the thesis, the kernel $p_k(s)$ in our model is set as a Gaussian power kernel. Indeed, other asymmetric form of $p_k(s)$ could be well studied for a specific data.

However, the choice for a good kernel can be very challenging. Observing the data could gives us, intuitively, some potential shapes for the kernel function. Suppose there is a set of L candidates $M = \{M_1, \dots, M_L\}$ under consideration for data \mathbf{x} . This is the problem of model selection, the goal is to find a model in M that most accurately represents the data according to some criterion of interest such as the Akaike Information criterion [104, 105], the Bayesian Information criterion [106, 107, 108]. In a Bayesian framework, Bayesian model comparison is a method of model selection based on Bayes factors [109]. By treating $p(M_l|\mathbf{x})$ as a measure of the “truth” of model M_l , a natural strategy is to choose the most probable M_l , i.e. the one for which $p(M_l|\mathbf{x})$ is largest.

The efficiency of the estimation is also a very important problem. The Metropolis-Hasting within Gibbs algorithm that has been developed for parameters estimation for our models provides interesting results. However, the drawback of MCMC methods is that it is difficult to assess when the Markov chain has reached its stationary state. Other potentially more robust and efficient population-based Monte Carlo algorithms have been established, including population-based MCMC [110] and sequential Monte Carlo sampler [111].

Under the authentication point of view, the impacts of the gray level distribution of the image to the performance of document authentication are substantial. Knowing the correct distribution can improve authentication performance because it is used for hypothesis testing and for channel optimization. Particularly, the distribution of the print-and-scan channel is required to calculate the likelihood ratio in likelihood ratio tests [12]. An accurate model can improve the performance of the likelihood ratio test. For the model of the shape of the microscopic dots, we had proposed an application in printers/printing technologies identification [27, 28]. The identification method can be extended using the new Markov spatial model.

Bibliography

- [1] “A serious problem for everyone,” http://ec.europa.eu/taxation_customs/customs/customs_controls/counterfeit_piracy/combating/index_en.htm, 2014.
- [2] R Bate, “The deadly world of fake medicine,” <http://edition.cnn.com/2012/07/17/health/living-well/falsified-medicine-bate/>, July 2012.
- [3] R Jotcham, “Overview of anti-counterfeiting technologies,” *The International anti-counterfeiting Directory*, pp. 29–32, 2005.
- [4] E. Mueller, P. Degott, and C.A. Despland, “Optically variable entity authenticating device and method,” Aug. 21 2014, US Patent App. 14/347,141.
- [5] M. Hein, M. Choate, N. Androff, and J. Gotro, “Visible and fluorescent dye containing laminate materials,” Feb. 14 2002, US Patent App. 09/928,103.
- [6] Baoshi Zhu, Jiankang Wu, and Mohan S. Kankanhalli, “Print signatures for document authentication,” in *Proceedings of the 10th ACM Conference on Computer and Communications Security*, New York, NY, USA, 2003, CCS ’03, pp. 145–154, ACM.
- [7] James DR Buchanan, Russell P Cowburn, Ana-Vanessa Jausovec, Dorothee Petit, Peter Seem, Gang Xiong, Del Atkinson, Kate Fenton, Dan A Allwood, and Matthew T Bryan, “Fingerprinting documents and packaging,” 2005.
- [8] William Clarkson, Tim Weyrich, Adam Finkelstein, Nadia Heninger, J Alex Halderman, and Edward W Felten, “Fingerprinting blank paper using commodity scanners,” in *Security and Privacy, 2009 30th IEEE Symposium on*. IEEE, 2009, pp. 301–314.
- [9] James D R Buchanan, Russell P Cowburn, Ana-Vanessa Jausovec, Dorothee Petit, Peter Seem, Gang Xiong, Del Atkinson, Kate Fenton, Dan a Allwood, and Matthew T Bryan, “Forgery: ‘fingerprinting’ documents and packaging,” *Nature*, vol. 436, no. 7050, pp. 475, July 2005.
- [10] Tobias Haist and Hans J Tiziani, “Optical detection of random features for high security applications,” *Optics communications*, vol. 147, no. 1, pp. 173–179, 1998.
- [11] “Authenticity and protection,” <http://www.crus-bourgeois.com/authentification-et-protection?lang=en>, April 2012.

Bibliography

- [12] Anh Thu Phan Ho, Bao An Mai Hoang, Wadih Sawaya, and Patrick Bas, “Document authentication using graphical codes: impacts of the channel model,” in *Proceedings of the first ACM workshop on Information hiding and multimedia security*. ACM, 2013, pp. 87–94.
- [13] Anh Thu Phan Ho, Bao An Hoang Mai, Wadih Sawaya, Patrick Bas, et al., “Authentication using graphical codes: Optimisation of the print and scan channels,” *EUSIPCO 2014*, 2014.
- [14] Anh Thu Phan Ho, Bao An Mai Hoang, Wadih Sawaya, and Patrick Bas, “Document authentication using graphical codes: Reliable performance analysis and channel optimization,” *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 9, 2014.
- [15] Bao An Hoang Mai, Wadih Sawaya, Patrick Bas, et al., “Image model and printed document authentication: A theoretical analysis,” in *IEEE International Conference on Image Processing*, 2014.
- [16] Luiz Velho and Jonas de Miranda Gomes, “Digital halftoning with space filling curves,” in *ACM SIGGRAPH Computer Graphics*. ACM, 1991, vol. 25, pp. 81–90.
- [17] Thrasyvoulos N Pappas and David L Neuhoﬀ, “Model-based halftoning,” in *Electronic Imaging’91, San Jose, CA*. International Society for Optics and Photonics, 1991, pp. 244–255.
- [18] Chen-Koung Dong, Thrasyvoulos N Pappas, and David L Neuhoﬀ, “Measurement of printer parameters for model-based halftoning,” in *IS&T/SPIE’s Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1993, pp. 355–366.
- [19] TN Pappas, “Digital Halftoning Techniques for Printing,” *Recent Progress in Digital Halftoning*, 1994.
- [20] Mark A Schulze and Thrasyvoulos N Pappas, “Blue noise and model-based halftoning,” in *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*. International Society for Optics and Photonics, 1994, pp. 182–194.
- [21] Thrasyvoulos N Pappas and David L Neuhoﬀ, “Printer models and error diffusion,” *Image Processing, IEEE Transactions on*, vol. 4, no. 1, pp. 66–80, 1995.
- [22] Thrasyvoulos N Pappas, “Model-based halftoning of color images,” *Image Processing, IEEE Transactions on*, vol. 6, no. 7, pp. 1014–1024, 1997.
- [23] Henry S Baird, “Document image defect models,” in *Structured Document Image Analysis*, pp. 546–556. Springer, 1992.
- [24] Elisa H Barney Smith, “Characterization of image degradation caused by scanning,” *Pattern Recognition Letters*, vol. 19, no. 13, pp. 1191–1197, 1998.

- [25] Elisa H Barney Smith, “Scanner parameter estimation using bilevel scans of star charts,” in *Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on*. IEEE, 2001, pp. 1164–1168.
- [26] Margaret Norris and Elisa H Barney Smith, “Printer modeling for document imaging,” in *Proceedings of the 2004 International Conference on Imaging Science, Systems, and Technology (CISST’04)*, 2004, pp. 1–7.
- [27] Q-T Nguyen, Yves Delignon, Lionel Chagas, and François Septier, “Printer technology authentication from micrometric scan of a single printed dot,” in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90280U–90280U.
- [28] Thong Q Nguyen, Yves Delignon, Lionel Chagas, and François Septier, “Printer identification from micro-metric scale printing,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6236–6239.
- [29] Helmut Kipphan, *Handbook of Print Media*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001.
- [30] Clyde J Fitch, “Electrophotographic printing-machine,” Sept. 24 1957, US Patent 2,807,233.
- [31] Julius R Insler, Allen R Leslie, and Ronald J Vigneri, “Laser printing system,” Aug. 7 1973, US Patent 3,751,587.
- [32] Kenji Aoki, Teiji Miura, Mitsuo Nagata, and Haruo Nakamura, “Electrophotographic printer,” June 7 1983, US Patent 4,386,836.
- [33] E Earle Thompson, “Electrophotographic printing subsystem with self-monitoring and adjustment,” Aug. 18 1998, US Patent 5,797,060.
- [34] John William Strutt and Lord Rayleigh, “On the instability of jets,” *Proc. London Math. Soc.*, vol. 10, pp. 4–13, 1878.
- [35] Henry S Baird, “The state of the art of document image degradation modelling,” in *Digital Document Processing*, pp. 261–279. Springer, 2007.
- [36] Robert Ulichney, *Digital halftoning*, MIT press, 1987.
- [37] Robert A Ulichney, “Review of halftoning techniques,” in *Electronic Imaging*. International Society for Optics and Photonics, 1999, pp. 378–391.
- [38] Tapas Kanungo, Robert M Haralick, and Ihsin Phillips, “Global and local document degradation models,” in *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*. IEEE, 1993, pp. 730–734.
- [39] Tapas Kanungo, “Document degradation models: Parameter estimation and model validation,” 1994.

Bibliography

- [40] Tapas Kanungo, Robert M Haralick, and Ihsin Phillips, “Nonlinear global and local document degradation models,” *International Journal of Imaging Systems and Technology*, vol. 5, no. 3, pp. 220–230, 1994.
- [41] Tapas Kanungo, Robert M. Haralick, Henry S. Baird, Werner Stuezel, and David Madigan, “A statistical, nonparametric methodology for document degradation model validation,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1209–1223, 2000.
- [42] Whitney K Newey and Daniel McFadden, “Large sample estimation and hypothesis testing,” *Handbook of econometrics*, vol. 4, pp. 2111–2245, 1994.
- [43] SJ Wright and J Nocedal, *Numerical optimization*, vol. 2, Springer New York, 1999.
- [44] Philip Wolfe, “Convergence conditions for ascent methods,” *SIAM review*, vol. 11, no. 2, pp. 226–235, 1969.
- [45] M. J. D. Powell, “A method for nonlinear constraints in minimization problems,” in *Optimization*, R. Fletcher, Ed., pp. 283–298. Academic Press, New York, 1969.
- [46] Magnus R Hestenes, “Multiplier and gradient methods,” *Journal of optimization theory and applications*, vol. 4, no. 5, pp. 303–320, 1969.
- [47] Pablo Pedregal, *Introduction to optimization*, Springer.
- [48] H. W. Kuhn and A. W. Tucker, “Nonlinear programming,” 1951.
- [49] C Radhakrishna Rao, “Information and accuracy attainable in the estimation of statistical parameters,” *Bulletin of the Calcutta Mathematical Society*, vol. 37, no. 3, pp. 81–91, 1945.
- [50] C. Gourieroux and A. Monfort, *Statistics and Econometric Models*, Number v. 1 in Statistics and Econometric Models 2 volume set. Cambridge University Press, 1995.
- [51] Harald Cramér, *Mathematical methods of statistics*, vol. 9, Princeton university press, 1999.
- [52] James O Berger, *Statistical decision theory and Bayesian analysis*, Springer, 1985.
- [53] Nicholas Metropolis and Stanislaw Ulam, “The monte carlo method,” *Journal of the American statistical association*, vol. 44, no. 247, pp. 335–341, 1949.
- [54] Christian P Robert and George Casella, *Monte Carlo statistical methods*, vol. 319, Citeseer, 2004.
- [55] Pranab K Sen and Julio M Singer, *Large sample methods in statistics: An introduction with applications*, vol. 25, CRC Press, 1994.

- [56] John Geweke, “Bayesian inference in econometric models using monte carlo integration,” *Econometrica: Journal of the Econometric Society*, pp. 1317–1339, 1989.
- [57] Walter R Gilks, Sylvia Richardson, and David J Spiegelhalter, “Introducing markov chain monte carlo,” in *Markov chain Monte Carlo in practice*, pp. 1–19. Springer, 1996.
- [58] Arnaud Doucet, Simon Godsill, and Christophe Andrieu, “On sequential monte carlo sampling methods for bayesian filtering,” *Statistics and computing*, vol. 10, no. 3, pp. 197–208, 2000.
- [59] Gerhard Winkler, *Image analysis, random fields and Markov chain Monte Carlo methods: a mathematical introduction*, vol. 27, Springer, 2003.
- [60] François Septier and Yves Delignon, “Mcmc sampling for joint estimation of phase distortions and transmitted symbols in ofdm systems,” *Digital Signal Processing*, vol. 21, no. 2, pp. 341–353, 2011.
- [61] James R Norris, *Markov chains*, Number 2008. Cambridge university press, 1998.
- [62] John Geweke et al., *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, vol. 196, Federal Reserve Bank of Minneapolis, Research Department, 1991.
- [63] Adrian E Raftery, Steven Lewis, et al., “How many iterations in the gibbs sampler,” *Bayesian statistics*, vol. 4, no. 2, pp. 763–773, 1992.
- [64] Andrew Gelman and Donald B Rubin, “Inference from iterative simulation using multiple sequences,” *Statistical science*, pp. 457–472, 1992.
- [65] Mary Kathryn Cowles and Bradley P Carlin, “Markov chain monte carlo convergence diagnostics: a comparative review,” *Journal of the American Statistical Association*, vol. 91, no. 434, pp. 883–904, 1996.
- [66] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller, “Equation of state calculations by fast computing machines,” *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [67] W Keith Hastings, “Monte carlo sampling methods using markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [68] Stuart Geman and Donald Geman, “Stochastic relaxation, gibbs distributions, and the bayesian restoration of images,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, , no. 6, pp. 721–741, 1984.
- [69] Peter Müller, “A generic approach to posterior integration and gibbs sampling,” 1991.

Bibliography

- [70] Luke Tierney, “Markov chains for exploring posterior distributions,” *the Annals of Statistics*, pp. 1701–1728, 1994.
- [71] Galin L Jones et al., “On the markov chain central limit theorem,” *Probability surveys*, vol. 1, no. 299-320, pp. 5–1, 2004.
- [72] Hugo Babij, Ramasamy Krishnan, Robert R Murray, Roland T Palmatier, and Marilyn C Yamat, “Offset lithographic printing process with a water based ink,” July 14 1998, US Patent 5,778,789.
- [73] Gaurav Sharma and Raja Bala, *Digital color imaging handbook*, CRC press, 2002.
- [74] Brian P Lawler, “Know thy enemy: understanding dot gain and its effects,” *Electronic Edition*, pp. 1–5, 1994.
- [75] Nobuyuki Otsu, “A threshold selection method from gray-level histograms,” *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.
- [76] Josef Kittler and John Illingworth, “Minimum error thresholding,” *Pattern recognition*, vol. 19, no. 1, pp. 41–47, 1986.
- [77] CH Li and Peter Kwong-Shun Tam, “An iterative algorithm for minimum cross entropy thresholding,” *Pattern Recognition Letters*, vol. 19, no. 8, pp. 771–776, 1998.
- [78] JN Kapur, Prasanna K Sahoo, and AKC Wong, “A new method for gray-level picture thresholding using the entropy of the histogram,” *Computer vision, graphics, and image processing*, vol. 29, no. 3, pp. 273–285, 1985.
- [79] Thrasyvoulos N Pappas, Jan P Allebach, and David L Neuhoff, “Model-based digital halftoning,” *Signal Processing Magazine, IEEE*, vol. 20, no. 4, pp. 14–27, 2003.
- [80] Takio Kurita, Nobuyuki Otsu, and N Abdelmalek, “Maximum likelihood thresholding based on population mixture models,” *Pattern Recognition*, vol. 25, no. 10, pp. 1231–1240, 1992.
- [81] Yakoub Bazi, Lorenzo Bruzzone, and Farid Melgani, “Image thresholding based on the em algorithm and the generalized gaussian distribution,” *Pattern Recognition*, vol. 40, no. 2, pp. 619–634, 2007.
- [82] Zhi-Kai Huang and Kwok-Wing Chau, “A new image thresholding method based on gaussian mixture model,” *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 899–907, 2008.
- [83] Cléo Baras and François Cayre, “Towards a realistic channel model for security analysis of authentication using graphical codes,” in *Information Forensics and Security (WIFS), 2013 IEEE International Workshop on*. IEEE, 2013, pp. 115–119.

- [84] Cléo Baras, François Cayre, et al., “Vers un modèle de canal réaliste pour l’analyse de la sécurité du processus d’authentification par code matriciel 2d,” *Proceedings GRETSI 2013*, pp. 1–4, 2013.
- [85] Lawrence T DeCarlo, “On the meaning and use of kurtosis.,” *Psychological methods*, vol. 2, no. 3, pp. 292, 1997.
- [86] Frank J Massey Jr, “The kolmogorov-smirnov test for goodness of fit,” *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
- [87] Norman L Johnson, Samuel Kotz, and N Balakrishnan, “14: Lognormal distributions,” *Continuous univariate distributions*, vol. 1, 1994.
- [88] Viktor Witkovský, “Computing the distribution of a linear combination of inverted gamma variables,” *Kybernetika*, vol. 37, no. 1, pp. 79–90, 2001.
- [89] John D Cook, “Inverse gamma distribution,” *online: http://www.johndcook.com/inverse_gamma.pdf*, Tech. Rep, 2008.
- [90] Norman L Johnson and Samuel Kotz, *Distributions in Statistics: Continuous Univariate Distributions: Vol.: 2*, Houghton Mifflin, 1970.
- [91] E Gómez, MA Gomez-Villegas, and JM Marin, “A multivariate generalization of the power exponential family of distributions,” *Communications in Statistics-Theory and Methods*, vol. 27, no. 3, pp. 589–600, 1998.
- [92] Miguel A Gómez-Villegas, Eusebio Gómez-Sánchez-Manzano, Paloma Maín, and Hilario Navarro, “The effect of non-normality in the power exponential distributions,” in *Modern Mathematical Tools and Techniques in Capturing Complexity*, pp. 119–129. Springer, 2011.
- [93] Solomon Kullback and Richard A Leibler, “On information and sufficiency,” *The Annals of Mathematical Statistics*, pp. 79–86, 1951.
- [94] David Arthur and Sergei Vassilvitskii, “k-means++: The advantages of careful seeding,” in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.
- [95] Raymond W Yeung, *A first course in information theory*, vol. 1, Springer, 2002.
- [96] I Csisz, Gábor Tusnády, et al., “Information geometry and alternating minimization procedures,” *Statistics and decisions*, 1984.
- [97] Urs Niesen, Devavrat Shah, and Gregory W Wornell, “Adaptive alternating minimization algorithms,” *Information Theory, IEEE Transactions on*, vol. 55, no. 3, pp. 1423–1429, 2009.
- [98] HE Daniels, “The asymptotic efficiency of a maximum likelihood estimator,” in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*. Univ of California Press, 1961, vol. 1, pp. 151–163.

Bibliography

- [99] David Hilbert, “Ueber die stetige abbildung einer line auf ein flächenstück,” *Mathematische Annalen*, vol. 38, no. 3, pp. 459–460, 1891.
- [100] Roger Fjortoft, Yves Delignon, Wojciech Pieczynski, Marc Sigelle, and Florence Tupin, “Unsupervised classification of radar images using hidden markov chains and hidden markov random fields,” *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 41, no. 3, pp. 675–686, 2003.
- [101] Noura Dridi, Yves Delignon, Wadih Sawaya, and Christelle Garnier, “Em-based joint symbol and blur estimation for 2d barcode,” in *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on*. IEEE, 2011, pp. 32–36.
- [102] Jian Zhang, Sei-ichiro Kamata, and Yoshifumi Ueshige, “A pseudo-hilbert scan algorithm for arbitrarily-sized rectangle region,” in *Advances in Machine Vision, Image Processing, and Pattern Analysis*, pp. 290–299. Springer, 2006.
- [103] Julian Besag, “Spatial interaction and the statistical analysis of lattice systems,” *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236, 1974.
- [104] Yosiyuki Sakamoto, Makio Ishiguro, and Genshiro Kitagawa, “Akaike information criterion statistics,” *Dordrecht, The Netherlands: D. Reidel*, 1986.
- [105] Olivier Alata and Ludovic Quintard, “Is there a best color space for color image characterization or representation based on multivariate gaussian mixture model?,” *Computer Vision and Image Understanding*, vol. 113, no. 8, pp. 867–877, 2009.
- [106] Gideon Schwarz et al., “Estimating the dimension of a model,” *The annals of statistics*, vol. 6, no. 2, pp. 461–464, 1978.
- [107] Olivier Alata and Christian Olivier, “Choice of a 2-d causal autoregressive texture model using information criteria,” *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1191–1201, 2003.
- [108] David Posada and Thomas R Buckley, “Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests,” *Systematic biology*, vol. 53, no. 5, pp. 793–808, 2004.
- [109] Robert E Kass and Adrian E Raftery, “Bayes factors,” *Journal of the american statistical association*, vol. 90, no. 430, pp. 773–795, 1995.
- [110] Ajay Jasra, David A Stephens, and Christopher C Holmes, “Population-based reversible jump markov chain monte carlo,” *Biometrika*, vol. 94, no. 4, pp. 787–807, 2007.
- [111] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra, “Sequential monte carlo samplers,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 3, pp. 411–436, 2006.

- [112] Norman L Johnson, Samuel Kotz, and N Balakrishnan, “Continuous univariate distributions , vol. 1john wiley & sons,” *New York*, 1994.
- [113] Edmund Taylor Whittaker and George Neville Watson, *A course of modern analysis*, Cambridge university press, 1927.

Appendix A.

Material

A.1. Parametric distribution

Some families of distributions have been constructed to provide approximations to as wide a variety of observed distributions as possible. The system of Pearson is one of these families, this is a four-parameters system [112]. For every member of the system, the probability density function $p(x)$ satisfies a differential equation of form

$$\frac{1}{p} \cdot \frac{dp}{dx} = -\frac{a+x}{c_0 + c_1x + c_2x^2}. \quad (\text{A.1})$$

Multiplying both sides by x^r

$$x^r(c_0 + c_1x + c_2x^2) \frac{dp(x)}{dx} + x^r(a+x)p(x) = 0,$$

assuming $x^r p(x) \rightarrow 0$ as $x \rightarrow \pm\infty$ for $r \leq 5$, then integrating both sides between $-\infty$ and $+\infty$, we obtain the equation

$$-rc_0\mu'_{r-1} + [-(r+1)c_1 + a]\mu'_r + [-(r+2)c_2 + 1]\mu'_{r+1} = 0 \quad (\text{A.2})$$

where μ'_r is the r -th moment, i.e. $\mu'_r = \int x^r p(x) dx$. Putting $r = 0, 1, 2, 3$ we gain the system of linear equations

$$\begin{pmatrix} -1 & 0 & 1 & 2\mu'_1 \\ -\mu'_1 & 1 & 2\mu'_1 & 3\mu'_2 \\ -\mu'_2 & 2\mu'_1 & 3\mu'_2 & 4\mu'_3 \\ -\mu'_3 & 3\mu'_2 & 4\mu'_3 & 5\mu'_4 \end{pmatrix} \begin{pmatrix} a \\ c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} \mu'_1 \\ \mu'_2 \\ \mu'_3 \\ \mu'_4 \end{pmatrix}. \quad (\text{A.3})$$

Solving this system we shall have the estimation of a, c_0, c_1, c_2 in term of moments.

Central moment μ_r is the moment about the mean μ , $\mu_r = \int (x - \mu)^r p(x) dx$. The random variable can always be arranged by the transform $X' = X - E[X]$ such that the expected value is zero, then $\mu'_1 = 0$ and $\mu'_r = \mu_r$ for $r \geq 2$. Then the formulas for a, c_0, c_1 and c_2 will be

$$c_0 = (4\beta_2 - 3\beta_1)(10\beta_2 - 12\beta_1 - 18)^{-1}\mu_2 \quad (\text{A.4})$$

$$a = c_1 = \sqrt{\beta_1}(\beta_2 + 3)(10\beta_2 - 12\beta_1 - 18)^{-1}\sqrt{\mu_2} \quad (\text{A.5})$$

$$c_2 = (2\beta_2 - 3\beta_1 - 6)(10\beta_2 - 12\beta_1 - 18)^{-1} \quad (\text{A.6})$$

Material

where $\gamma_1 = \sqrt{\beta_1}$ is index of *skewness*, $\gamma_1 = \mu_3(\mu_2)^{-3/2}$, and β_2 is index of *kurtosis*, $\beta_2 = \mu_4(\mu_2)^{-2}$.

The definition of the various types:

Type I : $\kappa = \frac{1}{4}c_1^2 (c_0c_2)^{-1} < 0$, a generalization of the beta distribution.

Type II : $\beta_1 = 0, \beta_2 < 3$, a special case of type I restricted to symmetric distributions.

Type III : $2\beta_2 - 3\beta_1 - 6 = 0$, gamma distribution belongs to this type.

Type IV : $0 < \kappa < 1$, Cauchy distribution belongs to this type.

Type V : $\kappa = 1$, inverse-gamma distribution is in this type.

Type VI : $\kappa > 1$, for example beta prime distribution is in this type.

Type VII : $\beta_1 = 0, \beta_2 > 3$, e.g. student's t-distribution.

By examining these parameters which are estimated from the observed data, we are able to predict choose a parametric distribution that fits well the observation.

A.2. Dot gain

Dot gain is a natural phenomenon in which the area of printed dots are growing larger than the theoretical image. For instance, the dot pattern cover 20% of the image, theoretically, but 30% area of the surface has been printed, then we say the dot gain is 10%. It can be expressed by a simple formula

$$a_g(\%) = a_{visual}(\%) - a_{theoretical}(\%)$$

where a_{visual} and $a_{theoretical}$ are, respectively, the visual area on the paper sheet and the area coverage of the theoretical one, a_g is the gained area. Dot gain is caused by ink spreading around the dot, called mechanical dot gain. Moreover, there is also an optical illusion creating gained area, Fig. A.1 shows one of these optical illusion.

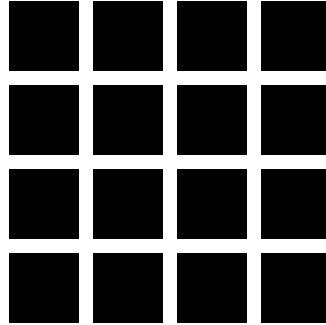


Figure A.1.: There are gray spots between the squares where they do not exist.

Appendix B.

Calculation in detail

B.1. Calculation of Fisher information (5.8)

$$\begin{aligned} \mathbb{E} \left[\nabla_{\theta} \nabla_{\theta}^T \ln p(\mathbf{U} \mid \theta) \right] &= \sum_{s \in S} \nabla_{\theta} \nabla_{\theta}^T \ln \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) \mathbb{E}(\delta[U_s]) \\ &\quad + \nabla_{\theta} \nabla_{\theta}^T \ln \prod_{k=1}^K (1 - p_k(s)) (1 - \mathbb{E}(\delta[U_s])) \end{aligned} \quad (\text{B.1})$$

we have

$$\begin{aligned} &\nabla_{\theta} \nabla_{\theta}^T \ln \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) \\ &= \nabla_{\theta} \frac{-\nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s)) \right)} \\ &= - \frac{\nabla_{\theta} \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s)) \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) + \nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s)) \right)^2} \end{aligned}$$

$$\begin{aligned} &\nabla_{\theta} \nabla_{\theta}^T \ln \prod_{k=1}^K (1 - p_k(s)) \\ &= \nabla_{\theta} \frac{\nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\prod_{k=1}^K (1 - p_k(s))} \\ &= \frac{\prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta} \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s)) - \nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(\prod_{k=1}^K (1 - p_k(s)) \right)^2} \end{aligned}$$

$$\mathbb{E}(\delta[U_s]) = P(U_s = 0 \mid \theta) = 1 - \prod_{k=1}^K (1 - p_k(s))$$

plug into (B.1), we obtain

$$\begin{aligned} &\mathbb{E} \left[\nabla_{\theta} \nabla_{\theta}^T \ln p(\mathbf{U} \mid \theta) \right] \\ &= - \frac{\nabla_{\theta} \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s)) \left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) + \nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s)) \right)^2} \\ &\quad + \frac{\nabla_{\theta} \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s)) \prod_{k=1}^K (1 - p_k(s)) - \nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\prod_{k=1}^K (1 - p_k(s))} \\ &= - \frac{\nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s)) \right)} - \frac{\nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\prod_{k=1}^K (1 - p_k(s))} \\ &= - \frac{\nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s)) \right) \prod_{k=1}^K (1 - p_k(s))}. \end{aligned}$$

Then the Fisher information matrix is

$$I(\theta) = -\mathbb{E} \left[\nabla_{\theta} \nabla_{\theta}^T \ln p(\mathbf{U} \mid \theta) \right] = \frac{\nabla_{\theta} \prod_{k=1}^K (1 - p_k(s)) \nabla_{\theta}^T \prod_{k=1}^K (1 - p_k(s))}{\left(1 - \prod_{k=1}^K (1 - p_k(s))\right) \prod_{k=1}^K (1 - p_k(s))}$$

B.2. Calculation of Fisher Information Matrix of model (5.17)

Recall the kernel

$$p(s) = \frac{n}{\pi \sigma^2 2^{\frac{1}{\beta}} \Gamma\left(1 + \frac{1}{\beta}\right)} \exp \left\{ -\frac{1}{2} \left(\frac{\|s - \mu\|_2^2}{\sigma^2} \right)^{\beta} \right\},$$

and the likelihood function of the model

$$q(\mathbf{u}) = \prod_s (1 - p(s))^{\mathbb{I}_{S_0}(s)} (1 - p(s))^{\mathbb{I}_{S_1}(s)}.$$

We have

$$\nabla_{\theta} p = \left(\frac{\partial p}{\partial \mu}, \frac{\partial p}{\partial \sigma^2}, \frac{\partial p}{\partial \beta}, \frac{\partial p}{\partial n} \right),$$

with

$$\begin{aligned} \frac{\partial p}{\partial \mu} &= p(s)(s - \mu) \frac{\beta \left(\frac{\|s - \mu\|_2^2}{\sigma^2} \right)^{\beta-1}}{\sigma^{2\beta}} \\ \frac{\partial p}{\partial \sigma^2} &= \frac{p(s)}{\sigma^2} \left[\frac{\beta}{2} \left(\frac{\|s - \mu\|_2^2}{\sigma^2} \right)^{\beta} - 1 \right] \\ \frac{\partial p}{\partial \beta} &= p(s) \left[\frac{\ln 2 + \psi\left(1 + \frac{1}{\beta}\right)}{\beta^2} - \frac{1}{2} \left(\frac{\|s - \mu\|_2^2}{\sigma^2} \right)^{\beta} \ln \left(\frac{\|s - \mu\|_2^2}{\sigma^2} \right) \right] \\ \frac{\partial p}{\partial n} &= \frac{p(s)}{n}. \end{aligned}$$

Plug into the formula (5.10)

$$I(\theta) = \sum_{s \in S} \frac{\nabla_{\theta} p(s) \nabla_{\theta}^T p(s)}{p(s)(1 - p(s))},$$

the elements of the matrix are

$$\begin{aligned}
 I_{\mu_1, \mu_2} &= \sum_{s \in S} r_s \left[\frac{\beta (\|s - \mu\|^2)^{\beta-1}}{\sigma^{2\beta}} \right]^2 (s_1 - \mu_1) (s_2 - \mu_2) \\
 I_{\mu_i, \sigma^2} &= \frac{1}{\sigma^2} \sum_{s \in S} r_s (s_i - \mu_i) \left[\frac{\beta}{2} \left(\frac{\|s - \mu\|^2}{\beta} \right)^\beta - 1 \right] \frac{\beta (\|s - \mu\|^2)^{\beta-1}}{\sigma^{2\beta}} \\
 I_{\mu_i, \beta} &= \sum_{s \in S} r_s (s_i - \mu_i) \frac{\beta (\|s - \mu\|^2)^{\beta-1}}{\sigma^{2\beta}} \left[\frac{\ln 2 + \psi \left(1 + \frac{1}{\beta} \right)}{\beta^2} - \frac{1}{2} \left(\frac{\|s - \mu\|^2}{\sigma^2} \right)^\beta \ln \left(\frac{\|s - \mu\|^2}{\sigma^2} \right) \right] \\
 I_{\mu_i, n} &= \frac{1}{n} \sum_{s \in S} r_s (s_i - \mu_i) \frac{\beta (\|s - \mu\|^2)^{\beta-1}}{\sigma^{2\beta}} \\
 I_{\sigma^2, \beta} &= \frac{1}{\sigma^2} \sum_{s \in S} r_s \left[\frac{\beta}{2} \left(\frac{\|s - \mu\|^2}{\sigma^2} \right)^\beta - 1 \right] \left[\frac{\ln 2 + \psi \left(1 + \frac{1}{\beta} \right)}{\beta^2} - \frac{1}{2} \left(\frac{\|s - \mu\|^2}{\sigma^2} \right)^\beta \ln \left(\frac{\|s - \mu\|^2}{\sigma^2} \right) \right] \\
 I_{\sigma^2, n} &= \frac{1}{n \sigma^2} \sum_{s \in S} r_s \left[\frac{\beta}{2} \left(\frac{\|s - \mu\|^2}{\sigma^2} \right)^\beta - 1 \right] \\
 I_{\beta, n} &= \frac{1}{n} \sum_{s \in S} r_s \left[\frac{\ln 2 + \psi \left(1 + \frac{1}{\beta} \right)}{\beta^2} - \frac{1}{2} \left(\frac{\|s - \mu\|^2}{\sigma^2} \right)^\beta \ln \left(\frac{\|s - \mu\|^2}{\sigma^2} \right) \right].
 \end{aligned}$$

with $i = 1, 2$ and $r_s = \frac{p_s}{1-p_s}$

B.3. Limit of the KL distance (5.19)

At first, there are some limits need to be discussed. To begin, with $x > 0$, we see that $\Gamma(x)$ approaches ∞ as $x \rightarrow 0$. In fact, since $e^{-t} \geq 1/e$ for $0 \leq t \leq 1$, one has

$$\Gamma(x) > \int_0^1 \frac{t^{x-1}}{e} dt = \frac{1}{ex}.$$

Because $\frac{1}{ex}$ goes to ∞ as $x \rightarrow 0$, $\Gamma(x)$ also tends to infinity as x approaches 0, $\lim_{x \rightarrow 0} \Gamma(x) = \infty$. Moreover, we are also interested in the limit of $\frac{\Gamma(\frac{1}{x})}{\Gamma(\frac{1}{x+\delta})}$ as $x \rightarrow \infty$, with fixed positive δ . Considering the Euler's formula for Gamma function (see [113] page 237)

$$\Gamma(x) = \frac{1}{x} \prod_{n=1}^{\infty} \left\{ \left(1 + \frac{1}{n} \right)^x \left(1 + \frac{x}{n} \right)^{-1} \right\}.$$

Then the ratio is rewritten as

$$\begin{aligned} \frac{\Gamma\left(\frac{1}{x}\right)}{\Gamma\left(\frac{1}{x+\delta}\right)} &= \frac{x}{x+\delta} \prod_{n=1}^{\infty} \frac{\left(1 + \frac{1}{n}\right)^{\frac{1}{x}} \left(1 + \frac{1}{(x+\delta)n}\right)}{\left(1 + \frac{1}{n}\right)^{\frac{1}{x+\delta}} \left(1 + \frac{1}{xn}\right)} \\ &= \frac{1}{1 + \frac{\delta}{x}} \prod_{n=1}^{\infty} \frac{\left(1 + \frac{1}{n}\right)^{\frac{\delta}{x(x+\delta)}} \left(1 + \frac{1}{(x+\delta)n}\right)}{\left(1 + \frac{1}{xn}\right)}, \end{aligned}$$

it is seen that limit of the ratio $\frac{\Gamma(\frac{1}{x})}{\Gamma(\frac{1}{x+\delta})}$ is 1 as $x \rightarrow \infty$. Recall the KL distance (5.19) with $\beta = \beta_0 + \delta$, δ is fixed and positive

$$\begin{aligned} d(\beta_0) &= \ln \frac{(\beta_0 + \delta) \Gamma\left(\frac{1}{\beta_0}\right)}{\beta_0 \Gamma\left(\frac{1}{\beta_0 + \delta}\right)} + \left(\frac{1}{\beta_0} - \frac{1}{\beta_0 + \delta}\right) \ln 2 + \frac{2^{\frac{\beta_0}{\beta_0 + \delta} - 1} \Gamma\left(\frac{\beta_0 + 1}{\beta_0 + \delta}\right)}{\Gamma\left(\frac{1}{\beta_0 + \delta}\right)} - \frac{1}{\beta_0 + \delta} \\ &= \ln \left(1 + \frac{\delta}{\beta_0}\right) + \ln \frac{\Gamma\left(\frac{1}{\beta_0}\right)}{\Gamma\left(\frac{1}{\beta_0 + \delta}\right)} + \left(\frac{1}{\beta_0} - \frac{1}{\beta_0 + \delta}\right) \ln 2 + \frac{2^{\frac{1}{1 + \frac{\delta}{\beta_0}} - 1} \Gamma\left(\frac{1 + \frac{1}{\beta_0}}{1 + \frac{\delta}{\beta_0}}\right)}{\Gamma\left(\frac{1}{\beta_0 + \delta}\right)} - \frac{1}{\beta_0 + \delta}. \end{aligned}$$

Combine with the above results, it is stated that $\lim_{\beta_0 \rightarrow \infty} d(\beta_0) = 0$.

B.4. Another real dots

Four printed dots with the resolution 600 dpi from HP-600M620 laser printer

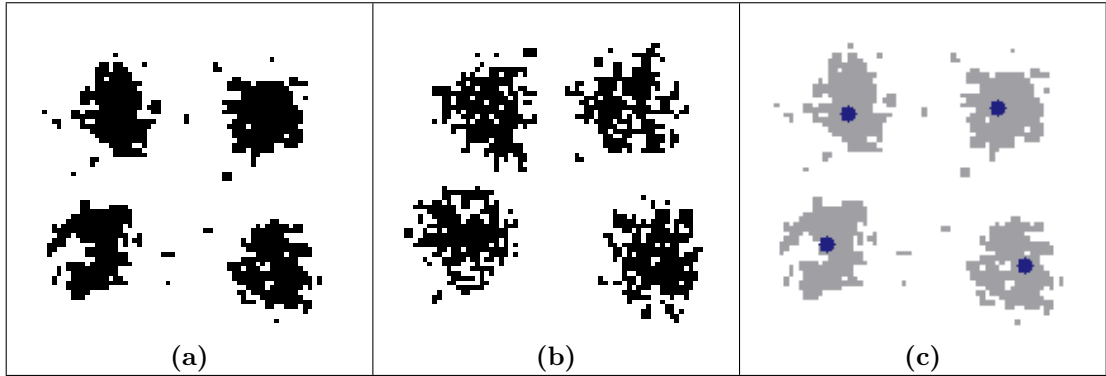


Figure B.1.: Real image vs. Simulated one, and the estimated centers.

	Estimators	Error with 99% of confidence
σ^2	66.82	± 0.019
β	1.36	$\pm 3.6 \times 10^{-4}$
η	0.76	$\pm 2.6 \times 10^{-4}$
λ	0.35	$\pm 1.2 \times 10^{-4}$

Table B.1.: Summary of the posterior samples mean of σ^2, β, η , and λ .